UNIVERSITY OF SOUTHAMPTON
Faculty of Engineering, Science and Mathematics
School of Geography

Understanding spatial uncertainty in empirical remote sensing models

by

Nicholas Alexander Samuel Hamm

Thesis submitted for the degree of
Doctor of Philosophy (Ph.D.)

March 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF GEOGRAPHY

Doctor of Philosophy

UNDERSTANDING SPATIAL UNCERTAINTY IN EMPIRICAL REMOTE SENSING MODELS

by Nicholas Alexander Samuel Hamm

This thesis addresses uncertainty in empirical remote sensing models. Specifically the empirical line method (ELM) for atmospheric correction of airborne remotely sensed data is examined.

First, the pairing of the field and remotely sensed data for input into the regression model is considered. The typical approach to the ELM averages over all field measurements in each ground target (GT). However, this approach is problematic since only a small number of data pairs are utilised in the regression modelling and this leads to a high degree of uncertainty in the regression model. The approach is also sensitive to sample size and to the use of sub-optimal GTs. Furthermore, the model is defined on a GT-sized support, whereas prediction is required on a pixel-sized support. These disadvantages were addressed by pairing the field measurements directly with the pixel-based remotely sensed data either using the point-pixel approach or block-pixel approach. The latter is favoured since it explicitly addresses the support issue, although both methods substantially reduce the uncertainty in the regression model. It is recommended that *at least* 50 and preferably 100 measurements should be obtained for each GT.

Utilisation of the point-pixel and block-pixel approach is contingent on accurate positioning of the field measurements both relatively and within the image. The impact of positional uncertainty is rarely considered in the literature and this thesis quantified the impact of positional uncertainty on the outcome of the ELM. When a moderate level of positional uncertainty was introduced, this led to bias in the parameter estimates for the point-pixel approach, although this could be minimised by using a sample size of *at least* 50 and preferably 100 measurements for each GT. For the geostatistical block-pixel approach introducing positional uncertainty led to an increase in the variogram at short lags but did not, ultimately, affect parameter estimation for the ELM.

Finally, adopting the point-pixel or block-pixel approach led to a regression model with heteroskedastic and spatially correlated residuals. However, these conditions are not handled in standard regression models. Hence a model that incorporates both weighting and spatial correlation was adopted. This was tested first with simulated data and found to yield unbiased estimates of all regression parameters, with the exception of the nugget variance. When this approach was applied to real data, it led to an increase in the uncertainty in the ELM. For the blue and green wavebands, this was manifested by an increase in the width of the confidence intervals. For the green and NIR wavebands, the predicted values were changed by up to 2% reflectance with confidence intervals in excess of 2% reflectance. Finally, it was demonstrated that spatial correlation may still be present in a random sample. Hence adopting a random sampling strategy does not obviate the need to model this phenomenon.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express particular thanks to my supervisors, Professor Peter Atkinson and Professor Ted Milton, for their encouragement, guidance and feedback over the years. I also acknowledge the support of the Natural Environment Research Council (NERC), who funded my studentship.

I would also like to thank Bill Damon and Karen Anderson for technical advice and assistance with field equipment and for fixing things when they went wrong. Thanks are also due to Charles Kerr and Jim Milne for being so helpful and accommodating with their IT support. The data providers also deserve acknowledgement. The NERC Airborne Remote Sensing Facility provided the remotely sensed data that was used in this thesis. The Ordnance Survey made the Landform PROFILE ® elevation data available to me prior to its incorporation into the Edina Digimap facility.

Thanks are also due to the staff and postgraduates in the School of Geography for providing a supportive and stimulating environment in which to work and study and for providing invaluable assistance with fieldwork. The people are too numerous to mention. However, I would particularly like to thank Stef Webb and Gary Llewellyn.

Finally I would like to thank my friends and family for their continued support and encouragement over the rather long gestation period of this thesis and for pushing me towards getting it finished. Special thanks are reserved for my parents and for Anna, for her continued tolerance whilst I wrote up the thesis.

List of abbreviations

Abbreviation	Definition
ARSF	Airborne Remote Sensing Facility (NERC)
ATM	Airborne Thematic Mapper
AVIRIS	Airborne Visible and Infra-Red Imaging Spectrometer
CASI	Compact Airborne Imaging Spectrometer
CNSD	Conditional Negative Semi Definite
CP	Coordinate-Positioning model
DEM	Digital Elevation Model
DGPS	Differential GPS
DTM	Digital Terrain Model
EGNOS	European Geostationary Navigation Overlay Service
EPSF	Equipment Pool for Field Spectroscopy (NERC)
ERS	European Remote sensing Satellite
ESA	European Space Agency
FOV	Field of View
FP	Feature-Positioning model
GCP	Ground Control Point
GIS	Geographic Information System
GPS	Global Positioning System
GT	Ground Target
IDS	Integrated Data System (NERC ARSF)
IFOV	Instantaneous Field of View
ILS	Incident Light Source (NERC ARSF)
JPL	Jet Propulsion Laboratory
KED	Kriging with an External Drift
KT	Kriging with a Trend
LiDAR	Light Detection And Ranging
LS	Least Squares
	cont/d

Abbreviation	Definition
MBG	Model Based Geostatistics
ME	Mean Error
MIMAS	Manchester InforMation and Associated Services
ML	Maximum Likelihood
MOD	Ministry of Defence (UK)
MSDR	Mean Square Deviation Ratio
MSE	Mean Square Error
NASA	National Aeronautics and Space Administration (USA)
NCAVEO	Network for Calibration and Validation for Earth Observation
NERC	Natural Environment Research Council (UK)
OK	Ordinary Kriging
OLS	Ordinary Least Squares
OS	Ordnance Survey
RAF	Royal Air Force (UK)
REML	Residual (or Restricted) Maximum Likelihood
RK	Regression Kriging
RMSRE	Root Mean Square Relative Error
RMSE	Root Mean Square Error
SAIL	Scattering by Arbitrarily Inclined Leaves
SAR	Synthetic Aperture Radar
SBAS	Satellite Based Augmentation System
SK	Simple Kriging
TM	Thematic Mapper (Landsat)
UCL	University College London
UK	United Kingdom of Great Britain and Northern Ireland
USA	United States of America
WLS	Weighted Least Squares

Chapter 1

Introduction

The aim of this thesis is to evaluate some of the major considerations that lead to uncertainty in empirical remote sensing models, with particular attention being given to the spatial domain. Specifically, the empirical line method (ELM) for atmospheric correction of airborne imagery was evaluated. The ELM is an empirical, regression-based technique. The issues addressed are also important when empirical methods are applied in other areas of remote sensing and geographic information science (GIS).

This chapter begins with a generic discussion of remote sensing models (section 1.1) and introduces the concept of uncertainty and related terminology (section 1.2). This is followed by a brief overview of the key issues that will be examined. The thesis structure is then outlined.

1.1 Remote sensing models

A remote sensing model may be represented by three linked subsystems: a *scene* model; an *atmospheric* model; and a *sensor* model (Strahler et al. 1986). The scene model describes the form and nature of the matter and electromagnetic energy within the scene, together with their spatial and temporal components. Scene models may be either discrete or continuous. In the former case, matter consists of discrete objects and energy fluxes. In the continuous case, energy and matter can be conceptualised as being continuous in time and space. The atmospheric model describes the interaction between the atmosphere and the electromagnetic energy entering and leaving the scene. The sensor model describes how the sensor responds to electromagnetic flux incident on it. The appropriate sensor model is dictated largely by engineering decisions made at the time of design and construction (Slater 1980). The output of the remote sensing system is an *image* model (Curran et al. 2000).

Operational use of remote sensing demands that inferences are made from image data. Thus, inferences of the properties and distribution of matter and energy within the scene

are made from the measurements that make up the image model. Such inference involves implementation of a remote sensing model. This will involve making assumptions about the scene, atmosphere and the sensor, even if explicit use, or reference to, these models is not made (Strahler et al. 1986).

Remote sensing models are not necessarily used in an operational context. Models may also be used as prognostic tools where exploring, characterising and understanding the relationship between remote sensing measurements and the scene is the aim (Verstraete et al. 1996).

From the above discussion, three purposes of remote sensing models may be established:

1. *Understanding*, exploring and characterising relationships between remotely sensed measurements and the scene.
2. *Prediction* or *estimation* of biological, chemical or physical entities from remotely sensed measurements.
3. *Integrating* remote sensing models and measurements with models of environmental processes. For example, remotely sensed data may be used to drive or update ecosystem or weather prediction models.

These three purposes are not mutually exclusive and, in many circumstances, there will be overlap among them. However, the last purpose lies outside the scope of the remote sensing model described above. It also lies outside the scope of the research conducted for this thesis. It is mentioned here for completeness since this activity is of great importance where remote sensing is applied in the environmental sciences such as glaciology (e.g., Bamber et al. 2000), ecosystems modelling (e.g., Wicks 2000, Shaw et al. 2000), hydrology (e.g., Bates et al. 1997, Burke et al. 1997), oceanography (e.g., Harmon & Challenor 1997, Subrahmanyam et al. 2001) and atmospheric science (e.g., Rodgers 2000).

Verstraete et al. (1996) distinguish between the properties of the scene that directly and indirectly influence the measurement. For example, the objective might be to predict biomass from radiance (the measurement). However, there is **not** a direct functional relationship between reflectance and biomass such that:

$$z = f(y) \tag{1.1}$$

where z is the measurement and y the variable of interest (in this case biomass).

The measurement is affected directly by the geometric arrangement of the media (i.e., size and orientation of items with which the radiation interacts), the physical, chemical and biological properties of the media (e.g., pigment concentration) and by the atmosphere. These properties can be represented by the set of state variables, $\mathbf{s}$. Therefore, at the most basic level:

$$z = f(\mathbf{s}) = f(s_1, \ldots, s_n) \tag{1.2}$$

where $\mathbf{s}$ is the set of state variables $(s_1, \ldots, s_n)$ needed to fully describe the physical state of the system. Thus, $f(\mathbf{s})$ models the functional relation between the scene (and atmospheric) components and the measurement and the variable of interest can be expressed as:

$$y = g(\mathbf{s}) \tag{1.3}$$

where g is a function that describes the relationship between the state variables and the variable of interest (y). Together, equations 1.2 and 1.3 demonstrate that there may not be a straightforward functional relationship between the variable of interest and the remotely sensed measurement.

The above formalism provides a useful framework on which to develop the discussion. Models can be divided into two classes: *deterministic* and *non–deterministic* models. The inverse of a deterministic model (represented by equation 1.2) may be expressed as:

$$s = f^{-1}(\mathbf{z}) \tag{1.4}$$

Most deterministic models are the physical models of the form discussed by Strahler et al. (1986) and Strahler (1994). These models have their roots in describing the physical interactions that occur. However, there are also deterministic empirical models where relationships are formed between variables (these may be formulated as in equation 1.1 or equation 1.2). The semi-empirical approach combines physical and empirical models (Strahler 1994). For example the bi-directional reflectance distribution function (BRDF) may be modelled as a weighted sum of a few empirical functions (the empirical component), where the functions are derived from physical approximations (the physical component) (Roujean et al. 1992).

By contrast, non-deterministic models provide an empirical relationship between variables. In this context the model is formulated on the basis of relationships between variables using appropriate statistical techniques. However, the relationship is not deterministic. Thus, if the objective is to predict biomass from remotely sensed data, the simple, first-order normal error regression model might be utilised:

$$y = \beta_0 + \beta_1 z + \varepsilon \tag{1.5}$$

where ε is an error term. This model is not deterministic, since there is no requirement for a functional dependence of y on z – rather z is used to account for the variance in y. β_0, β_1 and σ^2 are parameters which are estimated using an appropriate statistical technique (see chapter 2).

1.1.1 Estimation and prediction

Some matters of terminology need to be considered. This is important since different disciplines often attach different terminology to similar procedures or the same terminol-

ogy to different procedures. First, when adopting a statistical framework *parameters* and *variables* are clearly distinguished. For a given case parameters remain fixed or are drawn from a defined distribution (depending on whether a classical or Bayesian framework is adopted), whereas variables vary. This distinction is clear in equation 1.5. It is also clear in the model for a normal distribution, where the parameters are the mean and variance. This distinction often seems blurred in the literature on physical models. For example, leaf area index (LAI) might be described as a parameter, as is the case in the SAIL (Scattering by Arbitrarily Inclined Leaves) model of Verhoef (1984). This may not be appropriate, if we consider a statistical model of the type described in equation 1.5. However, the terms parameters and variables describe entities with respect to a given model. They do not describe the fundamental properties of the entity. In the SAIL model, LAI is fixed for a specific set of conditions. Thus, it is the viewing and illumination conditions that are the variables and it is consistent to regard LAI as a parameter. Under another set of conditions, LAI might be different, in the same way as β_0 and β_1 (equation 1.5) would change if the ambient conditions were changed. The value of the parameter may be determined by measurement, prior knowledge, or an estimation procedure.

It is also necessary to define prediction, estimation and retrieval. In this thesis, estimation shall be used where the aim is to derive the parameters of a model. The term prediction shall be used where a new value is to be derived on the basis of a model where the appropriate parameters and variables are known or have been estimated. The term retrieval is common in the model inversion literature, where the aim is to derive the model parameters from inversion of the forward model (Goel 1989, Kimes et al. 2000, Rodgers 2000). Although this is widely regarded as appropriate and legitimate, in this thesis the term prediction or estimation is used as appropriate, according to statistical convention.

1.2 Uncertainty

> "Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know." Donald Rumsfeld, US Secretary of Defense, February 12, 2002.

This statement won Secretary Rumsfeld the Plain English Campaign's 2003 "Foot in Mouth" campaign for a "baffling quote by a public figure"[1]; an award that, apparently, provoked outcries from academics and journalists. Indeed, it might be argued that Rumsfeld's statement covers the full range of uncertainty relevant to a range of disciplines including science, engineering, the social sciences, business and economics. However, he

[1]See: `http://www.plainenglish.co.uk/footinmouth.htm`

is guilty, perhaps, of placing crisp boundaries along a continuum ranging from complete understanding[2] to complete ignorance.

The concept of "known knowns" and "unknown unknowns" was developed (independently) by Professor Paul Curran in his welcome address to the Uncertainty in Remote Sensing and GIS conference held at the University at Southampton in July 2001 and developed in his foreword to the associated book (Curran 2002). Here, Curran distinguishes between "measurement uncertainty" and "understanding uncertainty". In this context the term measurement is not confined to raw data, but may also be the output of a model whereas "understanding uncertainty" refers to uncertainty in understanding natural, social or engineering phenomena. The key point is that, paradoxically perhaps, it is necessary to know something about uncertainty in order to quantify it. Indeed, it may be possible to be uncertain about the level of uncertainty in a measurement or model output.

This distinction between Curran's measurement uncertainty and understanding uncertainty was made several decades previously by Knight (1921) who draws his widely cited distinction between a measurable uncertainty (risk) and an unmeasureable one (uncertainty):

> "It will appear that a *measurable* uncertainty, or "risk" proper, as we shall use the term, is so far different from an *unmeasureable* one that it is not in effect an uncertainty at all. We shall accordingly restrict the term "uncertainty" to cases of the non-quantitative type. It is this "true" uncertainty, and not risk, as has been argued, which forms the basis of a valid theory for the divergence between actual and theoretical competition." p. 20.

This quotation, whilst widely cited, is at odds with the more common definition of risk, which is a function of uncertainty and the consequences of the (uncertain) event occurring (Adams 1995, Holton 2004). Since this thesis is concerned primarily with uncertainty, the issue of risk will not be considered further.

The discussion has avoided providing a simple, non-circular, definition of uncertainty. Indeed, further review of the literature would further expand the use of the term. Hence, the decision has been made to retreat to a dictionary definition of "not being definitely known or perfectly clear". Specific terminology, such as bias, precision and accuracy as well as confidence and prediction intervals, will be used to *quantify* particular aspects of uncertainty and to *qualify* what aspect of uncertainty is being considered.

Fisher (1999) considers three models of uncertainty for spatial data: error, vagueness and ambiguity. In this context, error may be attributed to deficiencies of the measuring device, natural variation and variation not accounted for by a model. This, Fisher (1999) asserts,

[2]Strictly this should state "slightly less than complete understanding" since scientific knowledge is *never* absolutely certain (Feynman 1988, May 2001).

can be accounted for using probabilistic approaches. Vagueness and ambiguity arise due to problems of classifying objects owing to the difficulty of making crisp distinctions (vagueness) or the difficulty of choosing between two or more alternatives (ambiguity). It is often argued that these forms of uncertainty are best dealt with by other, non-probabilistic approaches (e.g., Fisher 1999, Regan et al. 2002, Elith et al. 2002, Helton et al. 2004) – although this necessity is contradicted by O'Hagan & Oakley (2004)[3]. Ambiguity and vagueness are subsumed under the heading of *linguistic uncertainty* by Regan et al. (2002). Furthermore Foody (2003) distinguishes only two forms of uncertainty: ambiguity and vagueness and seems to subsume Fisher's (1999) error under the ambiguity heading.

For the purpose of this thesis, which is concerned with the modelling of continuous variables, the taxonomy of Fisher (1999) is most appropriate. Classification is not the topic of investigation, whereas quantification and, where possible, reduction of error is. Hence, a probabilistic statistical approach is adopted for this thesis.

In the context of the objectives of remote sensing models outlined in section 1.1 three forms of uncertainty require investigation (Chatfield 1995):

1. Uncertainty in the estimated parameters.
2. Uncertainty in the input variables and predicted variables.
3. Uncertainty in the model adopted. This may be termed *structural uncertainty* (Draper 1995) or *model inadequacy* (O'Hagan & Oakley 2004) and arises because the model is inadequate for the phenomena under investigation.

It will become clear that these three forms of uncertainty are interlinked. However, this categorisation is useful for discussion.

When considering empirical models of the type described in equation 1.5, uncertainty in the regression coefficients can be described using a confidence interval or probability distribution. The former would be adopted within the *frequentist* (or classical) statistical paradigm. Frequentists regard the model parameters as having fixed, but unknown values. Under their interpretation, an event's probability is defined as the limiting relative frequency in a large number of trials. Hence, a 95% confidence interval means that, if the experiment was repeated multiple times it would be expected that 95% of the intervals would contain the true parameter (Neter et al. 1996). Conversely, the Bayesian does not believe the parameter to have a fixed value and begins by assigning a prior distribution to the parameter, which is then updated using the model and the data (Box & Tiao 1992). The resulting probability distribution can then be interrogated to yield, for example, a 95% credible region. It may be noted that, if vague (uninformative) priors are used for

[3]It may be noted that O'Hagan and Oakley are prominent statisticians, whereas the other authors are drawn from mathematics, science and engineering. Hence, they are, perhaps, tackling the problem from different standpoints and, arguably, with different vested interests.

Bayesian estimation then the 95% credible interval will be the *same* as the 95% confidence interval. Despite the different interpretations, both provide a quantitative measure of uncertainty (Willink & Lira 2005).

Difference in the value of the confidence interval and credible interval will arise when informative prior information is built into the estimation phase. Furthermore, even if no prior information is available there are situations where the Bayesian approach offers greater flexibility. This is because models can be built *hierarchically* and uncertainty in the model output at one stage can be propagated to the next level in the hierarchy (Diggle et al. 1998, Wikle 2003). An example of such a model is the Bayesian model-based geostatistical approach outlined in section 2.2.2.5. Most of the analysis undertaken in this thesis adopts a frequentist framework and confidence intervals are calculated. More limited attention is given to the Bayesian approach.

Let us return to the use of confidence intervals or credible intervals as a measure of uncertainty in the estimated parameters. For a given model structure, the width of these intervals will be affected by the residual variance and the sample size. The residual variance is affected by measurement error in the data, natural variability in the data and model inadequacy. In the context of spatial data, it will also be affected by the support of the data and its positional uncertainty (Dungan 2002). Further discussion of these latter two factors is given in chapter 2 with analytical work presented in chapters 5 and 6. Measurement error and traceability to standards has a long history of investigation in the field of metrology (Barford 1985), and there are national and international standards for quantifying measurement error (e.g., Taylor & Kuyatt 1994). Nevertheless, such issues have received limited attention in remote sensing and the development of rigorous and traceable methodologies for calibration and validation of remotely sensed data is the objective of the UK Network for Calibration And Validation for Earth Observation[4] (NCAVEO) (Milton et al. 2004). This thesis focuses on issues surrounding spatial sampling and modelling, rather than instrument calibration. The other two sources of error given above were natural variability and model inadequacy. Regan et al. (2002) argue that natural variability could be eliminated if it could be modelled. However, this can be extraordinarily difficult and, in practice, it may be more efficient to model it as part of the error term (Armstrong 1998, Regan et al. 2002). Finally, in situations where the model can be linked to a functional process, it may be possible to validate the parameter estimates using independent data or the output from a physically based model.

When considering predicted variables, uncertainty can be quantified through a prediction interval or probability distribution. The width of the prediction intervals will be affected by the same factors that affect the confidence intervals as well as the values of the predictor variable(s). It is feasible to have narrow confidence intervals, but very wide prediction intervals. Such a situation would arise where the relationship between the predictor variables and the dependent variables is clearly defined, but the magnitude of the residuals

[4]NCAVEO: `www.ncaveo.ac.uk`

is large. In addition to interrogation of the prediction intervals the model residuals can be used to assess the validity of the model. For example, the residuals can be checked for independence and homoskedasticity. Furthermore, the predicted variable can be evaluated against a reference data set, such as an independent set of measured or simulated data.

The final area of uncertainty listed above was model uncertainty. This arises when the model is inadequate for the data and phenomenon under investigation. It should be noted that, even given an approximately perfect model, there will be some uncertainty remaining. This is referred to in the engineering and risk assessment literature as *aleatory uncertainty* (Oberkampf et al. 2004). Aleatory uncertainty can be thought of as inherent uncertainty that *cannot* be eliminated. Accordingly, there will always be some residual variance due to measurement error, natural variability, uncertainty over support and position and model inadequacy. In contrast *epistemic uncertainty* can be reduced by adoption of models that represent more accurately the process under investigation. Whilst it might be argued that, in the limit, *all* uncertainty is epistemic in practice the distinction is a useful one (O'Hagan & Oakley 2004).

There is an interesting twist to the above discussion on aleatory and epistemic uncertainty. A situation might be conceived where reducing epistemic uncertainty may actually lead to an increase in aleatory uncertainty. Indeed, this situation is demonstrated in chapter 7, where the regression model is developed to take account of residual variability (heteroskedasticity) and correlation. However, the resultant confidence and prediction limits are, in some cases, wider than the case where these factors were not modelled. In this situation increased understanding of the underlying process has led to an increase in the quantified uncertainty. In the framework of Rumsfeld, Curran and Knight, reducing ignorance[5] has led to an increase in the uncertainty estimated[6]. If it is accepted that the revised model is a more correct representation of the process then this increased level of uncertainty must also be accepted.

1.2.1 Accuracy, bias and precision

This section introduces three key terms that are used for uncertainty analysis: accuracy, bias and precision. It seems self-evident that something that is accurate should have low error (Taylor & Kuyatt 1994). More specifically, it can be regarded as a composite term which is a function of *bias* and *precision* (Atkinson 1999a, Atkinson & Foody 2002, Dungan 2002). Mathematically, this is given as (Dungan 2002) (see Casella & Berger

[5]Rumsfeld's *unknown unknowns*, Curran's *understanding uncertainty* and Knight's *uncertainty (true uncertainty)*

[6]Rumsfeld's *known unknowns*, Curran's *measurement uncertainty* and Knight's *risk (measurable uncertainty)*

(1990) p. 57 for details of the derivation):

$$\begin{aligned} MSE(\hat{y}) &= E[(\hat{y}-y)^2] \\ &= E[\{(\hat{y}-\bar{y})+(\bar{y}-y)\}^2] \\ &= E[(\hat{y}-\bar{y})^2]+E[(\bar{y}-y)^2] \\ &= Var[\hat{y}]+(\bar{y}-y)^2 \\ &= \texttt{imprecision} + \texttt{bias}^2 \\ &= \texttt{inaccuracy} \end{aligned} \tag{1.6}$$

where $\hat{y}$ is the predicted or measured value of the true value, y, and $\bar{y}$ is the sample average (expected value). Hence $E[(\hat{y}-\bar{y})^2] = Var[\hat{y}]$ is the variance of $\hat{y}$ about the average. In statistics, the inverse of the variance is termed the *precision*. The value $(\bar{y}-y)$ is the difference between the sample average and the true value and is termed the *bias*. Hence, accuracy is a function of bias and precision and for a measurement or model output to be accurate it must be both unbiased and precise. The level of accuracy required for something to be determined "accurate" is context dependent.

Dungan (2002) notes that some authors (e.g., Maling 1989, Goovaerts 1997, Mowrer & Congalton 2000) equate accuracy with unbiasedness. Indeed, this was the definition presented to me during my undergraduate degree. However, this notion is rejected by Dungan (2002) on the grounds that it ignores random variability. It is rejected in this thesis also, in favour of the definition presented in equation 1.6.

As discussed above, *precision* is given as the inverse of the sample variance, $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i-\bar{y}_i)^2 \tag{1.7}$$

Hence, as the variance decreases the precision increases. Alternatively, as the variance increases the *imprecision* also increases. The *bias* is given by the mean error:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i) \tag{1.8}$$

Finally, *accuracy* is quantified by the mean square error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{1.9}$$

the square root of which is commonly taken to yield the root mean square error (RMSE). These descriptive measures are useful and are widely used. However, they have their limitations. The ME and the $RMSE$ are not standardised, making it difficult to draw comparisons between predictions of variables with different means (Atkinson 1999a). This

can be addressed by providing the root mean square relative error (RMSRE), given as:

$$RMSRE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{y_i^2}} \tag{1.10}$$

A more fundamental limitation is that these measures provide single, global, values. Such global values may not be applicable across geographic space. A further problem relates to the difficulty of comparing predicted and reference data, and is discussed below.

When measurements (or simulations) are made of biophysical properties it is generally difficult to define a true value. This may be due to error in the instrument or measurement process or natural variation. Furthermore, different instruments and measurement techniques, that measure ostensibly the same entity, may actually measure different fundamental properties. This may be due to measurements being taken at different scales (Foody & Curran 1994b, Coops & Culvenor 2000), a problem which is apparent when soil hydraulic conductivity is measured (Paige & Hillel 1993, Reynolds 1994). It may also be noted that the entity commonly referred to as a measurement is often an abstraction, or result of a model (Reynolds & Zebchuk 1996). In remote sensing, radiance is often referred to as the measurement when, in fact, it is the output of the *sensor model* (Strahler et al. 1986) (See section 1.1). Voltage is the fundamental property that is measured and a sensor model, based on engineering considerations, is used to transform voltage to radiance (Slater 1980, Strahler et al. 1986). A further abstraction is required to transform radiance to reflectance. This discussion highlights two important factors:

1. Measurements are not fixed, known values (although, in certain circumstances it may be appropriate to regard them as such). Therefore, there is no fixed value to compare predicted (or estimated) values to. This is contrary to what is implied by equations 1.8 and 1.9.

2. When comparing variables (or parameters) predicted by a model to measurements, one needs to be sure that the same entity is being compared.

The above discussion illustrates the care that needs to be taken when implementing the methods for evaluating model accuracy using the ME and $RMSE$. This issue is given further consideration in chapter 8.

1.3 Aims and objectives

The ELM is a regression model, similar to that described by equation 1.5. Under the ELM, remotely sensed pixel-based measurements of radiance are the predictor variable and ground-based measurements of reflectance are the dependent variable. Although conceptually straightforward, problems arise to do with (*i*) standard practice, (*ii*) uncertainties in the data and (*iii*) violation of the requirements and assumptions of the regression

model. In particular, the spatial aspect of these issues is considered. In the first case, standard practice in the implementation of the regression models leads to large uncertainties in estimation and prediction. The reasons for this are discussed in section 3.2.3 and illustrated in section 5.1. The uncertainties in the data arise due to:

1. Different approaches to pairing the remotely sensed and field data.
2. Sample size and spatial sampling strategy.
3. Positional uncertainty in the location of the field measurements.
4. Spatial resolution of the remotely sensed imagery.

and the problems with the regression model arise due to:

1. Heteroskedasticity of the residuals.
2. The data are not independent.

The key issues to be addressed in this thesis are summarised as:

1. Pairing of the field and remotely sensed data for input into the ELM regression model. This includes consideration of sample size and examination of procedures for ensuring that the variables input into the ELM are defined on an appropriate spatial support.
2. Examining the effect of positional uncertainty in field data on results of the ELM regression model.
3. Addressing the issue of heteroskedasticity and spatial correlation in the residuals from empirical regression models.

In the context of the above issues the sample size and sample structure are also important. Sample structure refers to the layout of the sample. For example, data may be sampled at random locations in space or may be specified on a grid. In the latter case, there are various possible grid densities.

1.4 Thesis outline

This thesis is structured as follows:

1. Chapter 1 provides a generic introduction to remote sensing models and the concept of uncertainty. The thesis is also outlined.

2. Chapters 2 and 3 review the methodologies and techniques that will be used throughout the thesis. Chapter 2 covers aspatial and spatial statistical techniques and chapter 3 covers remote sensing with particular attention being given to atmospheric effects and atmospheric correction. Within this, the ELM is given specific attention.

3. Chapter 4 describes the field site, data collection, data pre-processing and experimental design. Summaries of the field data and image data are also presented.

4. Chapter 5 focuses on the approaches for pairing the remotely sensed and field data for input into the ELM. This begins with discussion and analysis of current standard practice. Approaches that explicitly seek to match the support of the remotely sensed and field data are then covered. The issue of sample size is also considered.

5. Chapter 6 focuses on uncertainties in the field data. Particular attention is given to positional uncertainty of the field data although consideration is also given to the interaction with sample size and pixel size.

6. Chapter 7 considers problems with the regression model. In particular, the regression model, as implemented in chapters 5 and 6, takes no account of heteroskedasticity or spatial autocorrelation in the residuals – thus violating standard regression model assumptions. Heteroskedasticity is addressed in this chapter.

7. Chapter 8 summarises the analysis and highlights the most significant findings, giving particular attention to advice for practitioners.

8. Chapter 9 provides a brief conclusion and suggests avenues for future research.

Chapter 2

Statistical models for remote sensing and GIS

Statistical models are used to model empirical relationships between variables. Such models are developed using statistical techniques and it should be recognised that where relations between variables are found this does not imply causal mechanisms. However, under well formulated experimental conditions statistical techniques may be used to investigate causal mechanisms (Neter et al. 1996). Furthermore, an understanding of causal mechanisms may be useful and important when developing predictive models.

Statistical techniques can be used to *explore* the relation between two or more variables. For example Atkinson & Emery (1999) used geostatistical techniques to explore the relation between spatial structure and reflectance at different wavelengths. They proposed using such analysis to inform the design of field sampling strategies. Statistical models, when combined with other theoretical and experimental work, may aid in the development of our understanding of causal mechanisms (Curran & Milton 1983, Neter et al. 1996).

Statistical models may also be used for *prediction.* Thus the value of the variable of interest can be predicted given one or more predictor variables and estimated values of the model parameters (as derived from an experimental data set). Of key importance here is the existence of a suite of graphical and statistical tests (Neter et al. 1996) that can be used to *(i)* determine the appropriateness of the statistical model, *(ii)* test the significance of the model parameters, *(iii)* determine confidence intervals about the estimated values of the parameters and *(iv)* determine prediction intervals about the predicted value. Furthermore, descriptive statistics may be used to compare predicted and measured values (Neter et al. 1996, Atkinson 1999a), where the measured values are not used in building the regression model. Most of the techniques described in this section are implemented in a *classical* framework, although Bayesian approaches are discussed briefly.

Much remote sensing research is based on developing reliable empirical statistical models that can be used to make predictions of the values of surface biophysical properties on

the basis of remotely sensed data (Verstraete et al. 1996). The data sets used in this type of work typically comprise field–based measurements of vegetation and soil properties and field spectroscopy data or airborne or satellite remotely sensed data (Curran & Hay 1986, Verstraete et al. 1996, Coops & Culvenor 2000). The remotely sensed data may be used as raw digital numbers (DN) (e.g., Trotter et al. 1997) or may be processed to radiance or reflectance (e.g. Salvador & Pons 1998). Often the remotely sensed data are processed further to form an *index* that is then used in the model instead of the original measurements. Examples include the Red Edge Position (REP) (e.g., Jago et al. 1999) and the Normalised Difference Vegetation Index (NDVI) (e.g., Larsson 1993). Note also that the field data are often processed prior to model development. For example, measurements of tree dimensions may be converted to a measure of biomass using empirically derived models available in the ecological literature (Curran & Hay 1986).

2.1 Regression models

In remote sensing, empirical statistical models are generally based on regression, or some development of it, although correlation models are also used. In this section, simple linear regression will be introduced first and more complex models will then be introduced. In particular, mechanisms for accounting for correlation and heteroskedasticity in the residuals will be considered. The discussion will then be expanded to include the incorporation of spatial data.

2.1.1 Linear regression

Regression analysis serve three major purposes: *(i)* description and exploration, *(ii)* control and *(iii)* prediction (Neter et al. 1996). Exploration and prediction were discussed above. In the case of control, an objective might be to determine whether a new measurement lies within prescribed confidence limits, specified on the basis of a computed regression model (Neter et al. 1996).

In the simple case where there is one predictor variable, the model can be written as (Neter et al. 1996):

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{2.1}$$

where:

$i = 1, \ldots, n$,

y_i is the value of the response variable in the ith trial,

x_i is the predictor variable in the ith trial and is a known constant,

β_0 and β_1 are parameters (also known as coefficients). Note that, whilst β_1 is the coefficient of x, β_0 can be thought of as the coefficient of a dummy variable which always takes the value 1. These coefficients can be thought of as the *"y–intercept"* (β_0) and *"slope"* (β_1),

ε_i is the error (also known as the residual) and is a random variable. The ε_is are independent and identically distributed (iid) (i.e., the error terms are not correlated), with mean $E[\varepsilon_i] = 0$ and variance $Var[\varepsilon_i] = \sigma^2$ for all i.

Furthermore, it can be shown that:

$$Var[y_i] = Var[\varepsilon_i] = \sigma^2 \tag{2.2}$$

and that $Var[y]$ is the same for all levels of x. Note also that any two responses, y_i and y_j are uncorrelated. Detailed discussion and proof is provided by Neter et al. (1996: p11). Note that this proof is predicated on the case that x_i is a known quantity (i.e., $Var[x_i] = 0$) and, therefore, β_1 and ε_i are independent.

The model given in equation 2.1 is a *simple, first-order model.* It is simple because it contains only one predictor variable and is first-order because it is linear in the predictor variables and linear in the parameters. Mathematically, it is straightforward to extend this to cover multiple predictor variables, although choosing which variables to include is more challenging (Neter et al. 1996, Ver Hoef et al. 2001). Non-linear models are more complex, but can also be developed (Sen & Srivastava 1990, Neter et al. 1996).

For a sample of (x_i, y_i) pairs, point estimators of β_0 and β_1 may be derived using the method of *least-squares (LS).* The method of least squares proceeds by minimising the sum of square errors (SS_E), given as

$$SS_E = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2 \tag{2.3}$$

Hence, the process of LS estimation can be conceptualised as adjusting the values of β_0 and β_1 to minimise the SS_E. This calculation may be performed using numerical search procedures. However, for many models (including the model given in equation 2.1), an analytical solution is available, as:

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \tag{2.4}$$

$$b_0 = \frac{1}{n}\left(\sum y_i - b_1 \sum x_i\right) = \bar{y} - b_1 \bar{x} \tag{2.5}$$

where the LS estimators of β_0 and β_1 are denoted b_0 and b_1 respectively. Note that it is also commonplace to use the circumflex to indicate an estimator (e.g., $\hat{\beta}_0$ and $\hat{\beta}_1$). This notation is also used in this thesis. By reference to the *Gauss-Markov theorem* it follows that the least squares estimators are *unbiased* and have *minimum variance* among all

unbiased linear estimators (Neter et al. 1996, p20). Hence:

$$E[b_0] = \beta_0 \tag{2.6}$$

$$E[b_1] = \beta_1 \tag{2.7}$$

$$\hat{y}_i = b_0 + b_1 x_i \tag{2.8}$$

where $\hat{y}_i$ is the point estimator of y_i, and:

$$y_i = b_0 + b_1 x_i + e_i \tag{2.9}$$

where e_i is the residual. Using least squares, the variance (σ^2) given in equation (2.2) is estimated as:

$$s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - p} = \frac{\sum e_i^2}{n - p} \tag{2.10}$$

where s^2 is the estimator of σ^2 and p is the number of model coefficients (two in the simple, first order model).

The model given by equation 2.1 is limited, since it is not possible to set up interval estimates or make tests on b_0, b_1, $\hat{y}_i$ or e_i. To achieve this, an assumption needs to be made about the distribution of ε_i. When the form of the probability distribution of the error terms (σ^2) is known, estimators of the parameters β_0, β_1 and σ^2 can be obtained by the *maximum-likelihood (ML)* method. Maximum likelihood estimation is very important and is discussed in more detail in section 2.1.3.

The standard assumption that is made is that the error terms, ε_i, are normally distributed. This gives rise to the *normal error model*:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{2.11}$$

which is as equation 2.1 but *with the condition that the* $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$ and all other terms are as defined for equation 2.1. It should also be noted that, since $Var[y_i] = Var[\varepsilon_i] = \sigma^2$ (from equation 2.2), the y_i are also drawn from a normal distribution. Under the normal error model the estimators of β_0 and β_1 retain the properties that they are unbiased and have minimum variance amongst all unbiased linear estimators.

In reality, data never strictly conform to a normal distribution. However, in practice, this is found to be useful (Box & Tiao 1992, Neter et al. 1996). There are a range of factors that contribute to the residual variance and these tend to combine to form a normal distribution. Furthermore, hypothesis testing procedures are based on the t-distribution and tend only to be sensitive to large departures from normality.

The normal error model provides a framework for evaluation of the appropriateness of a model and for establishing uncertainty in the model. In this context, diagnostic plots are highly valuable. However, it is also possible to set up interval estimates or make tests

on b_0, b_1, $\hat{y}_i$ or e_i. This is important since it provides a mechanism for examining the appropriateness of the model and for evaluating uncertainty in the model parameters and predictions based on the model.

The variance of the b_0 and b_1 ($Var[b_0]$ and $Var[b_1]$) are estimated as $s^2_{b_0}$ and $s^2_{b_1}$ respectively (for formulation see Neter et al. 1996). This makes it possible to test whether the parameters are significantly different from 0 using the t–test, for a user-selected confidence level. It also allows confidence intervals to be set, using the t distribution, whereby it can be shown that:

$$b_1 - [t(1-\alpha/2; n-p)]s_{b_1} \leq \beta_1 \leq b_1 + [t(1-\alpha/2; n-p)]s_{b_1} \tag{2.12}$$

where $1-\alpha$ is the user defined confidence level (e.g., at the 95% confidence level $1-\alpha = 0.95$) and $n-p$ specifies the degrees of freedom (p is the number of parameters). The same procedure may be followed for b_0.

The confidence band for the entire regression line can also be calculated. Following from this, prediction intervals for y_h given a new value of x_h can be determined using:

$$\hat{y}_h - [t(1-\alpha/2; n-p)]s_{pred} \leq y_h \leq \hat{y}_h + [t(1-\alpha/2; n-p)]s_{pred} \tag{2.13}$$

$$s^2_{pred} = s^2 + s^2_{\hat{y}_h} \tag{2.14}$$

Note that s_{pred} incorporates two components:

1. $Var[y_h]$ (i.e. σ^2), which is estimated as s^2 (see equation (2.10)). This is the variance of the distribution of y at $x = x_h$. This can be thought of as variability arising from the spread of the residuals.

2. $Var[\hat{y}_h]$, which is estimated as $s^2_{\hat{y}_h}$. This is the variance in the sampling distribution of $\hat{y}_h$. This arises owing to variability caused by uncertainty in the slope and intercept of the regression line.

In addition to the above statistics the *coefficient of determination (R^2)* is widely used. This is a *descriptive* measure of the association between x and y in the regression model. It is defined as follows:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{2.15}$$

here, SS denotes sum of squares. SS_T is the total sum of squares, SS_R is associated with the regression and SS_E is associated with the residuals (error), they are related as follows:

$$SS_R = \sum (\hat{y}_i - \bar{y})^2 \tag{2.16}$$

$$SS_E = \sum (y_i - \hat{y}_i)^2 \tag{2.17}$$

$$\begin{aligned} SS_T &= \sum (y_i - \bar{y}_i)^2 \\ &= SS_R + SS_E \\ &= \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2 \end{aligned} \tag{2.18}$$

Hence, R^2 provides a measure of the reduction of the total variation gained by use of the predictor variable and reflects the proportion of SS_T which is partitioned into SS_R. In the case of a perfect relation $SS_E = 0$ and $R^2 = 1$. The *coefficient of correlation (R)* is given by $\sqrt{R^2}$.

It should be noted that R^2 is simply a descriptive measure. It is **not** a formal test for the regression model and does not allow assessment of confidence intervals. This descriptor should not be used in isolation, but may be used in conjunction with other techniques when assessing utility of a regression model. Neter et al. (1996) note that R and R^2 are widely used but are subject to misuse and poor interpretation. In particular, a large R^2 does not necessarily indicate that accurate predictions can be made or that the regression line is a good fit and a low R^2 does not necessarily indicate that the variables are not related.

2.1.2 Matrix notation

The regression model can also be expressed using matrix notation. This style is widely used and preferred by some authors. Indeed, matrix notation is more succinct and often more straightforward to interpret.

In matrix form, equation 2.1 may be expressed as:

$$\mathbf{y} = \beta\mathbf{X} + \varepsilon \tag{2.19}$$

where $\beta = (\beta_0, \beta_1)^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, $\mathbf{y} = (y_1, \ldots, y_n)^T$ and $\mathbf{X}$ is an $n \times 2$ matrix where the first column is a column of 1s and the second column is $(x_1, \ldots, x_n)^T$.

2.1.2.1 Estimating the regression parameters

Using matrix notation LS- and ML-estimated regression coefficients are given as:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{2.20}$$

which is algebraically equivalent to calculating the results of equations 2.4 and 2.5.

The covariance matrix of $\mathbf{b}$ is then given as:

$$\begin{aligned} C(\mathbf{b}) &= s^2(\mathbf{X}^T\mathbf{X})^{-1} \\ &= \begin{bmatrix} Var(b_0) & Cov(b_1, b_0) \\ Cov(b_0, b_1) & Var(b_1) \end{bmatrix} \end{aligned} \tag{2.21}$$

2.1.3 Maximum-likelihood estimation

Maximum-likelihood estimation (MLE) is of great importance in statistical inference. To understand this, it is useful to begin by considering a probability density function (pdf) such as the normal distribution, which is given as:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \tag{2.22}$$

where μ and σ^2 are the mean and variance parameters respectively and x is some variable of interest. If μ and σ^2 are known, then the probability density $p(\cdot)$ can be obtained for any given value of x. However, if there are several realisations of x but μ and σ^2 are unknown then some method of estimating these parameters is required. This is achieved by recasting $p(x)$ as $L(\mu, \sigma^2|\mathbf{x})$. Whereas $p(x)$ can be thought of as a function of x where μ and σ^2 are fixed, $L(\mu, \sigma^2|\mathbf{x})$ can be thought of as a function of μ and σ^2 where $\mathbf{x} = (x_1, \ldots, x_n)^T$ is fixed. The objective of MLE is to establish the parameter values that are most consistent with the sample data. This is done by maximising the likelihood function:

$$L(\mu, \sigma^2|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right) \tag{2.23}$$

In the context of the normal error regression model μ is replaced with $\hat{y}_i = \beta_0 + \beta_1 x_i$, where y is the response variable and x now denotes the predictor variable . The likelihood is written as:

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2|\mathbf{y}) &= \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2\right) \end{aligned} \tag{2.24}$$

which is expressed in matrix form as:

$$L(\beta, \sigma^2|\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)\right) \tag{2.25}$$

where $\beta = (\beta_0, \beta_1)^T$

This approach is conceptually similar to least squares in that both LS and MLE both work by optimising a function. When adopting LS, the SS_E is minimised whereas with MLE the likelihood function is maximised. However, an important difference is that ML estimation requires making an explicit distributional assumption – in this case the normal error model. As with LS, the ML function can be maximised numerically although, in many cases, analytical solutions exist. Under the normal error model the ML estimators of β_0 and β_1 are the same as the LS estimators (see equations 2.4, 2.5 and 2.21). However,

the ML estimator of σ^2 is given as:

$$\hat{\sigma}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n} = \frac{\sum e_i^2}{n} \tag{2.26}$$

Note that the LS estimator (equation 2.10) compensated for the loss of degrees of freedom associated with estimating the regression coefficients (β_0 and β_1) by using $n - p$ in the denominator. The ML estimator does not make this compensation and is known to yield a biased estimate. However, for large samples, this bias should be small. Nevertheless, the LS estimator of the variance is often preferred for this reason. Alternatively, the ML estimator can be adjusted for the degrees of freedom (Harris & Johnson 1996)

For the types of model discussed so far ML estimation offers no practical benefit over LS estimation. However, ML does offer greater flexibility in more complex circumstances. For example, it does not restrict the user to working with normal distributions. Importantly, it also allows relaxation of the assumption that $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, such that $\varepsilon_i \sim MVN(0, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a covariance matrix and MVN refers to the multi-variate normal distribution. The likelihood is then given as:

$$L(\beta, \theta|\mathbf{Y}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\beta)\right) \tag{2.27}$$

where θ is a vector of parameters that parameterise $\mathbf{\Sigma}$. This relaxation of the iid requirement is useful since the residuals from a regression model may exhibit heteroskedasticity (i.e., non-constant error variance) and/or a correlation structure. This is not straightforward and is discussed in detail in sections 2.1.4, 2.2.2.4 and 2.4.

Finally, it is often more convenient to work with log-likelihoods since this may simplify the algebra. This applies particularly to numerical maximisation and is used in subsequent analysis sections. Hence, the log-likelihood of equation 2.27 is given as:

$$l(\beta, \theta|\mathbf{y}) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\beta) \tag{2.28}$$

2.1.4 Developing the simple linear model

In many cases the simple linear model may prove to be inadequate. Most commonly, it is necessary to include more than one covariate to model variability in y, such that

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_{p-1} x_{i,p-1} + \varepsilon_i \\ &= \beta_0 + \sum_{k=1}^{p-1} \beta_k x_{i,k} + \varepsilon_i \end{aligned} \tag{2.29}$$

where p is the number of model coefficients and $p-1$ is the number of predictor variables. Models with two or more predictor variables can be expressed in matrix form using the same format as for models with one parameter (equation 2.20).

Further modifications to the regression model may be made to account for *non-linear* relationships and non-normally distributed residuals. This lies outside of the scope of this thesis, although discussion can be found in Neter et al. (1996) and Dobson (1990).

In the context of this thesis, the following three issues are of particular interest.

1. The error terms do not have constant variance (i.e., $Var[\varepsilon_i] \neq \sigma^2]$ for all i). Hence, the residuals are heteroskedastic;
2. The error terms are not independent.
3. There is error in x.

A common approach to the first situation is to weight the observations according the residual variance (σ_i^2) for that observation (i). If the σ_i^2s are known this is straight forward. In practice, these are rarely known. Nevertheless, if the σ_i^2s are known up to a constant of proportionality, the σ_i^2 can be written as (Sen & Srivastava 1990, Neter et al. 1996, Gelman et al. 2004):

$$\sigma_i^2 = \frac{1}{w_i}\sigma^2 \tag{2.30}$$

where the i denotes a particular observation and σ_i^2 is the variance of the error term associated with the ith observation. The w_i are weights attached to each observation and σ^2 is the constant of proportionality which needs estimating. The "weights" are so called because observations with a high variance are "down weighted' in the parameter estimation (Sen & Srivastava 1990, Neter et al. 1996, Gelman et al. 2004). If the relative magnitudes of the σ_i^2 are known, these can be used to provide the weights (w_i) in equation 2.30. The weights then form the elements of a diagonal matrix:

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \tag{2.31}$$

This matrix allows generalisation from $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, such that $\varepsilon_i \sim MVN(0, \sigma^2\mathbf{W})$. The regression coefficients are then estimated as:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y} \tag{2.32}$$

which is often known as *weighted-least-squares* (WLS) (Sen & Srivastava 1990, Neter et al. 1996, Upton & Cook 2002). Next, σ^2 is estimated as:

$$s^2 = \frac{\sum w_i(y_i - \hat{y}_i)^2}{n - p} \tag{2.33}$$

Finally, the covariance matrix of $\hat{\beta}$ is given as

$$Cov(\hat{\beta}) = s^2(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1} \tag{2.34}$$

If the w_i are not known, they need to be estimated. There are various possibilities available here:

1. Neter et al. (1996) provide various suggestions for predicting the w_i by modelling the residuals, e_i, as a function of predictor variables x_i. This is based on the fact that e_i^2 is an estimator of σ_i^2 (and $|e_i|$ and estimator of σ_i). For example, one of the following strategies might be adopted:

$$\hat{s}_i = \beta_0 + \beta_1 x_i \tag{2.35}$$

$$\hat{s}_i^2 = \beta_0 + \beta_1 x_i \tag{2.36}$$

   The weights are then given as the inverse of s_i^2, i.e., $w_i = 1/s_i^2$ and σ^2 is modelled as the s^2. Note that the x_i above could be replaced by another variable, including $\hat{y}_i$ or a variable not included in the regression model. In order to obtain the e_i, one might begin with the residuals obtained from OLS and then iterate to obtain the new set of residuals.

2. Neter et al. (1996) also note that, where replicate measurements are made at different levels of the predictor variable, these replicates can be used to estimate the weights. If a sufficiently large number of replicates have been obtained, then the weights can be estimated directly. Alternatively, the sample variances or standard deviations can be regressed against the appropriate predictor variable, as described above. For experiments, replicates can be included in the experimental design although for observational studies, near replicates may be used. A similar approach is suggested by Gelman et al. (2004), who advocate dividing the data set into batches of similar variance.

3. Finally, it could be assumed that the weights are directly related to a third variable and treated as if they are known.

The WLS approach is straight forward to implement and is covered in many statistical texts (e.g., Sen & Srivastava 1990, Gelman et al. 1995, Neter et al. 1996). In practice, specification of the weights is a difficult task, and is glossed over in such texts. This is addressed in chapter 7.

The second case refers to non-independence in the error terms with respect to x. Such a situation may manifest itself in the form of non-constant error variance . Neter et al. (1996) discuss non-independence of the error term when the data are collected in a sequence (such as time or space). Therefore, they suggest plotting residuals against the sequence index

(e.g. time, time order, distance). This procedure may highlight correlation between the error terms and space or time. It should be realised that autocorrelation along x is strictly different to autocorrelation along space or time. It may be appropriate to view temporal (or spatial) dependency as a missing variable which could be incorporated into the regression via equation 2.29 (Augustin et al. 1996). A better solution may be to adopt a model which explicitly deals with temporal (Neter et al. 1996) or spatial autocorrelation (Oliver & Webster 1986, Atkinson 1999a, Atkinson 1999b, Fotheringham et al. 2000, Lark 2000). If the spatial structure amongst the residuals can be established, then it is possible to generalise $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, such that $\varepsilon_i \sim MVN(0, \mathbf{\Sigma})$. Following from this, equations 2.32 and 2.34 can be generalised to

$$\hat{\beta} = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{y} \tag{2.37}$$

which is often known as *generalised-least-squares* (GLS) (Sen & Srivastava 1990, Neter et al. 1996, Upton & Cook 2002) and the covariance matrix of $\hat{\beta}$ is given as

$$Cov(\hat{\beta}) = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \tag{2.38}$$

The covariance matrix, $\mathbf{\Sigma}$ looks like:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & 0 & \cdots & \sigma_n^2 \end{bmatrix} \tag{2.39}$$

where each off-diagonal element describes the covariance between two specified data-points. The challenge in this case is establishing the structure of $\mathbf{\Sigma}$. Modelling of spatial structure will be discussed in detail in section 2.2.

In addition to the cases described above, there are two further cases which are likely to be of particular interest in remote sensing. First, the situation where there is error or uncertainty in the measurement of x_i presents an important barrier to applying regression models in remote sensing (Curran & Hay 1986). The problem arises owing to the basic requirement that the error terms be independent of the predictor variables (Sprent 1969, Kendall & Stuart 1967, Curran & Hay 1986, Webster 1989, Neter et al. 1996). If the error in x_i is denoted η_i, we have:

$$\eta_i = x_i - x_i^* \tag{2.40}$$

where x_i is the measured value of the variable and x_i^* the true value. Combining equations 2.40 and 2.41 yields:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i^* + \varepsilon_i \\ &= \beta_0 + \beta_1 (x_i - \eta_i) + \varepsilon_i \\ &= \beta_0 + \beta_1 x_i + (\varepsilon_i - \beta_1 \eta_i) \end{aligned} \tag{2.41}$$

and it is clear that x_i^* is not independent of the error term $(\varepsilon_i - \beta_1\delta_i)$ (see Neter et al. (1996, p165) for proof) .

Dealing with the problem of measurement error in x is a large topic (Sprent 1969, Kendall & Stuart 1967, Curran & Hay 1986, Fuller 1987, Webster 1989, Neter et al. 1996, Fernandes & Leblanc 2005) which is dealt with briefly here. Sprent (1969) and Kendall & Stuart (1967) begin by postulating a *functional* relationship between x and x

$$y_i^* = \beta_0 + \beta_1 x_i^* \tag{2.42}$$

where y_i^* and x_i^* are not observable, but the random variables Y_i and X_i are:

$$x_i = x_i^* + \eta_i \tag{2.43}$$

$$y_i = y_i^* + \varepsilon_i \tag{2.44}$$

They then impose the following restrictions:

$$\begin{aligned} \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \text{ and } \eta_i \sim N(0, \sigma_\eta^2), \text{ for all } i; \\ Cov(\varepsilon_i, \varepsilon_j) = Cov(\eta_i, \eta_j) = 0,\ i \neq j; \\ Cov(\varepsilon_i, \eta_j) = 0, \text{ for all } i,\ j. \end{aligned} \tag{2.45}$$

If the ratio $\lambda = \sigma_{\varepsilon^2}/\sigma_{\eta^2}$ is known, then β_1 can be estimated using ML as:

$$\hat{\beta}_1 = \frac{\sum y_i^2 - \lambda \sum x_i^2 + \sqrt{\{(\sum y_i^2 - \lambda \sum x_i^2)^2 + 4\lambda(\sum x_i y_i)^2\}}}{2\sum x_i y_i} \tag{2.46}$$

where $\lambda = \sigma_\varepsilon^2/\sigma_\eta^2$. Note that this formulation is achieved by setting the origin to $\overline{x}, \overline{y}$ (i.e., β_0 is set to zero).

The unbiased estimator of σ_ε^2, is then given as

$$\hat{\sigma}_\varepsilon^2 = \frac{\lambda}{(n-2)(\lambda + \hat{\beta}_1^2)}(\sum y_i^2 - 2\hat{\beta}_1 \sum x_i y_i + \hat{\beta}_1^2 \sum x_i^2) \tag{2.47}$$

Furthermore, procedures similar to those outlined in section 2.1.1 can be used to test whether $\beta_0 \neq 0$ and $\beta_1 \neq 0$ and to set confidence intervals and prediction intervals. Further details can be found in Kendall & Stuart (1967).

If $\lambda = 1$, this is equivalent to minimising the variance perpendicular to the regression line which may be termed the *principal axis* (Webster 1989). Note that this differs from the oft-cited reduced major axis (RMA) (Webster 1989) approach. Under the RMA approach $\hat{\beta}_1 = \sigma_Y/\sigma_X$ and the regression line always bisects the angle between the regression of y on x and x on y (Kermack & Haldane 1950, Curran & Hay 1986). The RMA is dependent on the relative magnitudes of the standard deviations of x and y, whereas the principal axis is dependent on the relative magnitudes of the *errors* in x and y.

The above issue is important in remote sensing, since regression is often used under conditions where there is measurement error in x (Curran & Hay 1986). For example, a regression relation may be developed between reflectance and biomass. However, *the process of deriving* reflectance and biomass introduces uncertainty into these quantities. Some researchers (e.g. Curran & Hay 1986, Larsson 1993) have aimed to deal with this problem by using modified regression procedures taken from the statistical literature (e.g. Fuller 1987). Salvador (1999) adopted a novel procedure for dealing with cases where there is uncertainty in the position of the field measurements in relation to the remote sensing measurement. This approach is based on incorporating variance in biomass into the predictive model.

Second, there may be situations when prediction of $\hat{x}$ for a measured value of y is required (e.g., Curran & Hay 1986, Trotter et al. 1997). Neter et al. (1996) provide a methodology for predicting $\hat{x}$ and deriving confidence limits. However they also note that, *"the inverse prediction problem has aroused controversy amongst statisticians"* (p169), with some suggesting that inverse regression should be performed (i.e. x should be regressed on y).

2.2 Spatial analysis

As with regression modelling, spatial analysis may be performed with the objective of *(i)* describing, exploring and understanding a data set or *(ii)* prediction – a process that is often referred to as spatial interpolation. Indeed, it is usual to undertake *(i)* before *(ii)*. Remote sensing provides spatially referenced data. However, the spatial component of the data is often not utilised. The spatial component may be of great use for enhancing the information content in remotely sensed data (e.g., Ni et al. 1999, Lewis et al. 1999, Llewellyn et al. 2000), especially where the objective is to understand how the scene structure affects the remotely sensed measurement. The spatial component has also been used for designing efficient strategies for field sampling (e.g., Atkinson & Emery 1999), where the objective is to characterise the spatial variation on the ground, both at the small scale (e.g., at the sub-pixel level) and gaining wide enough coverage at the larger scale (e.g., to ensure that enough pixels are sampled). Finally, the spatial component may be of value for predicting some variable, where a remotely sensed image (and possibly some field data) provides the predictor variables (Atkinson et al. 1992, Atkinson et al. 1994, Ashton 1998, Atkinson 2000).

Spatial analysis is based on the notion that all places are similar, but nearby places are more similar than distant places. This notion is attributed to Waldo Tobler and assigned the status of Tobler's First Law of Geography (Longley et al. 2005). The term *spatial structure* is used to describe the form of spatial association and may be quantified by the degree of *spatial autocorrelation*. The term spatial autocorrelation refers to correlation amongst instances of the same variable (hence "auto") realised at different locations (hence

"spatial") (Ver Hoef et al. 2001).

There are two basic factors that may give rise to spatial autocorrelation (Augustin et al. 1998). First, there may be *apparent* autocorrelation, whereby predictor variables that are not included in the regression model (they may not even have been measured) covary with space (or time). Second, there may be *true* spatial correlation. An example of the latter case might be created by a gregarious animal that is present at a particular location not only because of the suitable habitat, but because others of its species are also present (Augustin et al. 1998). In other circumstances, it may be convenient and appropriate to treat apparent spatial (or temporal) autocorrelation as true autocorrelation. For example, in the study presented by Atkinson (2000) it is intuitively sensible to assume that flooded locations will occur adjacent to other flooded locations, so it is reasonable to argue that it demonstrates "true" spatial correlation. It may be possible to address the problem of autocorrelation by including a measure of location (e.g. Eastings and Northings) as a covariate (e.g., Buckland & Elston 1993), by adding further predictor variables, or by adopting a method that deals explicitly with temporal or spatial autocorrelation (Augustin et al. 1996, Neter et al. 1996).

Procedures for spatial analysis may be divided along four general categories (Burrough & McDonnell 1998, Heywood et al. 2006):

1. *Local or global.* Global methods of interpolation use all the data in the region of interest to derive a model and provide predictions for the whole area. Local methods operate in a small defined area around the point (or area) of interest. This notion is developed in more detail below.

2. *Exact or approximate.* Exact interpolators honour the data (i.e., they predict the measured value at locations where a measurement has been taken. Approximate interpolators simply predict based on the chosen model and estimated parameters. This might be visualised as the difference between the measured value and the regression line. These different approaches are partly a characteristic of the model chosen and partly a function of purpose. For example, trend surfaces are approximate interpolators and tend to produce smooth surfaces, whereas kriging is typically an exact interpolator. Further discussion with respect to kriging is given in sections 2.2.2.3 and 2.2.2.5.

3. *Gradual or abrupt.* These approaches are distinguished by the continuity of the surface, although this should not be a binary classification.

4. *Deterministic or statistical.* Deterministic models give a prediction that is a mathematical or geometrical manipulation of the input data. Statistical models incorporate an explicitly probabilistic framework and allow a measure of uncertainty to be attached to each prediction. Statistical models are of primary interest in this thesis.

There are a variety of local models for spatial analysis. One approach would be to conduct a series of local analyses and then examine how various diagnostic features vary across the area of interest. This allows local regression analyses to be performed, and variation in parameter estimates across the scene can subsequently be examined and used for prediction. This forms the basis of geographically weighted regression (GWR) (Fotheringham et al. 2000, Fotheringham et al. 2002). Other local methods include inverse distance weighting (IDW), splines and nearest neighbour methods (Burrough & McDonnell 1998). Nearest neighbour methods include Thiessen polygons and Delaunay triangulation. A key attraction of local methods is that they account explicitly for local variation and hence may be used to account for non-stationarity. The concept of stationarity is discussed in section 2.2.2.1.

Global models include *trend-surfaces* (Burrough & McDonnell 1998) and *spatially autoregressive models* (Atkinson 2000). Trend surface models are, essentially, regression models that use the location (Eastings and Northings) as covariates. These may be augmented or replaced by further covariates or an autocovariate (producing a spatially autoregressive model). The term "autocovariate" refers to the predictor variable (or some transformation of it) that has been measured or predicted at another location.

In geostatistics, the spatial covariance (one variable) or spatial cross–covariance (more than one variable) is modelled explicitly. Thus, it is possible to model spatial variation in a variable using the *variogram* or covariance function and to predict the value of the variable of interest at points where it is not sampled, by interpolation using the technique of *kriging* (one variable) or *cokriging* (more than one variable) (Isaaks & Srivistava 1989, Webster & Oliver 2001). A typical application of cokriging would be the prediction of some variable of interest on the ground (such as biomass), where that variable is sparsely sampled but where some other variable is intensively sampled (e.g. an image covering the area of interest is available) (Atkinson et al. 1992, Atkinson et al. 1994). Geostatistics was originally developed in the mining industry (Matheron 1963, Journel & Huijbregts 1978, Armstrong 1998) although it has also found wide application in the environmental sciences (Cressie 1993, Soares et al. 1996, Webster & Oliver 2001) including remote sensing (Curran & Atkinson 1998, Lark 2000) and geographical information science (Burrough & McDonnell 1998, Lark 2000, Longley et al. 2005). Geostatistics offers an advantage over conventional regression techniques since it incorporates explicitly the spatial component.

The basic form of geostatistical and spatially autoregressive models is limited, since it does not account for local variation in the autocorrelation structure. The local analysis methods (such as GWR) do deal with local variation in the parameter values. However, they do not deal explicitly with autocorrelation or with cross–correlation. Spatially autoregressive models and geostatistical models tend to assume stationarity across the region of interest and hence constant autocorrelation structure.

There is some vagueness as to whether geostatistical models are local or global. Geosta-

tistical models are *estimated* using the entire data set and make some form of stationarity assumption (see section 2.2.2) so are clearly global models in these respects. However, when it comes to *prediction* (kriging) (see section 2.2.2.3) the model is typically applied in a restricted area around the point or block of interest. Furthermore, although the entire data set is used for model estimation, the maximum lag is often restricted by the user, rather than being the maximum lag in the data. In these senses, kriging might be said to be a local interpolation technique. This issue is explored in more detail in sections 2.2.2.3 and 2.2.2.5.

2.2.1 Scale and spatial support

Scale is an important issue in remote sensing as well as GIS and other areas of spatial modelling. It impacts on a broad range of application areas including basic remote sensing analysis (Aman et al. 1992, Raffy 1994), ecology (Jelinski & Wu 1996, Van Gardingen et al. 1997, Dungan et al. 2002), hydrology (Stewart et al. 1996) and human geography (Fotheringham & Wong 1991, Martin 1996, Cockings & Martin 2005). However, the meaning of scale is multifaceted and merits some brief discussion here. Montello (2001) notes that there are three domains of scale: spatial, temporal and thematic (or attribute). In this section, focus will be given to spatial scale, although it will be shown that these are related. Dungan et al. (2002) conceptualises scale along three axes: phenomenon scale, scale of sampling (also measurement or observation) and analysis scale. It is feasible that the measurements may be obtained at a different scale from those required for analysis and may also be different from the scale of the phenomenon under investigation. Under such scenarios, careful consideration needs to be given to matching the scales (Foody & Curran 1994a, Heuvelink & Pebesma 1999, Montello 2001, Dungan 2002, Dorren & Heuvelink 2004).

The term scale may be ascribed several different meanings (Dungan et al. 2002). A geographer will often think of scale in terms of the *cartographic ratio* which describes ratio of a distance on a map to that distance in reality. Hence, small (cartographic)-scale maps cover large areas and large-scale maps cover small areas. However, this definition is at odds with the more commonly understood notion of the word which tends to associate small-scale with small things and large-scale with large things (Foody & Curran 1994a, Longley et al. 2005) – a notion that is, more precisely, ascribed to the word *extent*. The term *grain* is often used to describe some property of the phenomenon or observations (e.g., the minimal elementary sampling unit) (Dungan et al. 2002). However, Dungan et al. (2002) notes a lack of consensus over the precise meaning of the word grain and recommends that it be substituted for a less ambiguous term or, at least, be strictly defined in a given context. The term *resolution* refers to the minimum unit that can be resolved and applies to the observations and analysis, rather than phenomena (Dungan et al. 2002). Finally, Dungan et al. (2002) provides the terms *lag* and *support*. Lag refers to the spatial

separation between unit phenomena or neighbouring sampling or analysis units. Support refers to the size and shape of the measurement unit. In this thesis, the concept of the support is most relevant and is discussed in more detail below.

The support refers to the volume of the measurement of the property of interest (Journel & Huijbregts 1978, Armstrong 1998). For optical remote sensing applications it is sufficient to think of the support as referring to the area rather than the volume, since there is no depth component (Dungan 2002). If measurements are made over two different supports using identical measurement techniques (i.e., the support of the measurement is the only component that varies), then the statistical properties of the two measurements will differ (Journel & Huijbregts 1978, Armstrong 1998). Consider two sets of measurements, one with small support, v, and the other with large support, V, which are acquired within a region, R. The *dispersion variance* of v within R is written $D^2(v, R)$. It will be found that $D^2(V, R) < D^2(v, R)$. Higher order moments (such as skewness and kurtosis) may also change (Journel & Huijbregts 1978). Therefore, the histogram of measurements on support v will differ from that on support V, although the arithmetic mean will be the same. Note that, at the limits, the small support v might be a point and the large support V might be the region, R.

The situation may arise where data are available on support v, but are required on support V. This is the so-called change-of-support problem (COSP) (Cressie 1996). Where v is exhaustive (i.e., a census) then the change-of-support problem is relatively straightforward, since the mean (or some other property, such as the median) can readily be calculated. However, although the calculations are straightforward the implications of both the level of aggregation (i.e., the size of support V) and the zonation (i.e., the placement of the boundaries) are important. This is the so-called modifiable areal unit problem (MAUP), which is discussed below. Where v is a sample, it is necessary to interpolate onto a surface in order to be able to aggregate to support V. This is discussed in section 2.2.1.1.

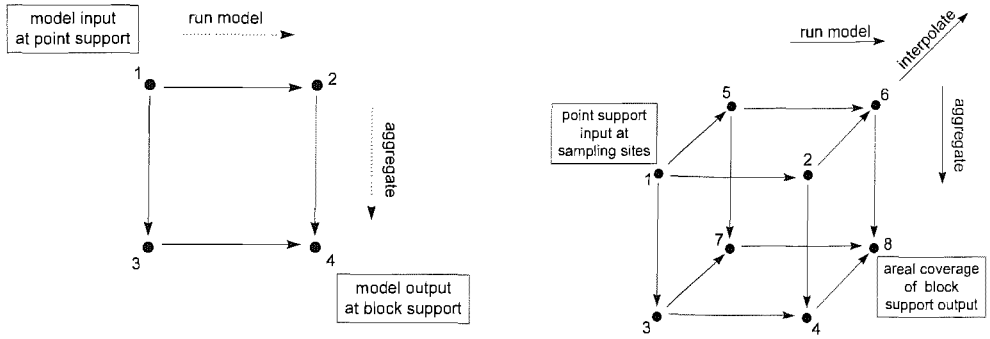
This change of support from v to V is often referred to as *scaling up* or *upscaling* (e.g., Stewart et al. 1996, Van Gardingen et al. 1997, Bierkens et al. 2000). Similarly, the change of support from V to v is often termed *downscaling*. However, Heuvelink & Pebesma (1999) are careful to distinguish between the term upscaling and aggregation. They apply the term aggregation to *data* and the process of aggregation is independent of any model that may subsequently be used. In contrast, they use the term upscaling to refer to the case where *processes* (or parameters representing processes) are upscaled from support v to support V. They cite the example of hydraulic conductivity, which is often measured over small soil volumes, such as sediment cores. However, such measurements cannot straightforwardly be aggregated to V for use in regional groundwater models since the parameter takes on a different meaning in the context of such a model. In this context, the method for upscaling is defined in the context of the model that will be used (Heuvelink & Pebesma 1999), and may be substantially more complex than straightforward aggregation (Wen & Gomez-Hernandez 1996). In this thesis, the terms aggregation (disaggregation) and upscaling (downscaling) will be applied according to the

definition given by (Heuvelink & Pebesma 1999), rather than using them interchangeably.

Consideration of the support is important when different variables are combined in a model (Heuvelink & Pebesma 1999, Bierkens et al. 2000, Dungan 2002, Heuvelink 2002, Dorren & Heuvelink 2004). This is because a model that is developed for variables that are defined on one support may not be valid when the variables are defined on a different support. A further issue arises when a model is developed and parameterised on support v but the user requires the model output (i.e., predictions based on the model) on support V. This point is discussed in detail by Heuvelink & Pebesma (1999) who propose two alternatives (figure 2.1(a)). The first is to aggregate the input variables from support v to support V and then to run the model ($1 \rightarrow 2 \rightarrow 4$). The second is to run the model first and then to aggregate ($1 \rightarrow 3 \rightarrow 4$). This choice is important, because it can, ultimately, lead to different model outcomes. For a deterministic, error-free, linear model both routes will produce the same outcome. However, the outcomes will be different for non-linear models. However, where there is uncertainty in the model (or the data) different outcomes may be observed for both linear and non-linear models. The choice of which route to take depends largely on the support on which the model was developed and calibrated. However, many process models are based on theory which applies at the point support. Hence, for such models, Heuvelink & Pebesma (1999) recommend running the model *before* aggregating ($1 \rightarrow 3 \rightarrow 4$). However, it may be noted that this route may not always be possible. For example, data may not be available at the point support or running the model at this scale may be prohibitively expensive. In such cases, an assessment of the impact of this choice is required (Aman et al. 1992). Dungan (2002) notes that the support issue has received limited attention in remote sensing.

It is important to recognise that the situation posed in figure 2.1(a) applies only to the case where exhaustive spatial data are available. However, where field (rather than remotely sensed) data are obtained, the data set will, almost always, not give exhaustive spatial coverage. Hence an *interpolation* step is required to move from a spatial sample to a surface. This can be conceptualised as a cube (figure 2.1(b)) and the objective is to progress from location 1 to 8 in this cube. Since it is not possible to aggregate *before* interpolating, it is not possible to move through (3) or (4). This leaves three possible routes: $1 \rightarrow 2 \rightarrow 6 \rightarrow 8$, $1 \rightarrow 5 \rightarrow 6 \rightarrow 8$ or $1 \rightarrow 5 \rightarrow 7 \rightarrow 8$. The key difference between the first two pathways is whether interpolation should be performed before or after running the model. Heuvelink & Pebesma (1999) advocate interpolating *before* running the model. Their rationale is that the model output is a function of the various input data sets. Hence, by interpolating the model output less information is being exploited than if the model inputs are interpolated. Furthermore, for a spatial model that uses neighbouring values, it will be necessary to interpolate before running the model. This leaves a choice between the second and third pathways. The choice here will be dependent on whether the model is defined at the point or block support.

The above discussion draws attention to a closely related issue, known as the modifiable areal unit problem (MAUP). The MAUP relates to the *aggregation* and *zonation* mecha-



(a) Exhaustive spatial coverage

(b) Incomplete spatial coverage

FIGURE 2.1: Possible routes from point-support data to block-support model output. Figure 2.1(a) is for exhaustive spatial coverage, whereas figure 2.1(b) gives the options where spatial coverage is incomplete and interpolation is required. Source: Heuvelink & Pebesma (1999)

nism employed when changing from smaller to larger supports (Dungan et al. 2002, Longley et al. 2005). This is illustrated in figures 2.2(a), 2.2(b), 2.2(c), 2.2(d). As mentioned above aggregating the small blocks will not change the regional mean, but may change the median as well as the histograms and higher order moments. It may also change the spatial patterns that are observed. A similar effect may be observed when different zonation structures are implemented. This is demonstrated by the summary statistics presented in table 2.1.

As well as affecting the univariate properties, the MAUP also has implications for multivariate modelling. Notably, multivariate models that are defined at support v may not be valid at support V. Furthermore, significantly different empirical relationships may be observed for different levels of aggregation and for different zonations at the same level of aggregation (Fotheringham & Wong 1991, Jelinski & Wu 1996). Indeed, Fotheringham & Wong (1991) find their results *"rather depressing in that they provide strong evidence of the unreliability of any multivariate analysis undertaken with data from areal units"* (p1025). Cressie (1996) demonstrates that, in theory, the MAUP can be addressed by modelling explicitly the "geography". However, he also notes that, in practice, the mechanism for modelling geography is not clear. A phenomenon that is related to the MAUP is the *ecological fallacy* which states that predictions made from models built with aggregated data may not be accurate for individuals (Martin 1996). Indeed, this concept applies more generally and has been termed the "cross-level fallacy" (Montello 2001).

The MAUP is most often associated with GIS (Longley et al. 2005) and particularly with socio-economic applications (e.g., Fotheringham & Wong 1991, Martin 1996, Cockings & Martin 2005). However, its relevance in environmental sciences is also recognised (Jelinski & Wu 1996, Journel 1996a, Dungan et al. 2002). Research on the MAUP dates back several decades (e.g., Gehlke & Biehl 1934, Openshaw & Taylor 1979), but continues to be an active and challenging research area (Cressie 1996, Dungan 2002, Baveye 2004), even though the term MAUP is not always used (e.g., Aman et al. 1992, Raffy 1994, Heuvelink &

Figure	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Mean	Standard Deviation
2.2(a)	2	79	150	260	735	201	166
2.2(b)	44	102	163	265	505	201	129
2.2(c)	75	107	186	256	476	201	111
2.2(d)	49	109	216	297	363	201	107
2.2(e)	2	93	195	253	735	194	171
2.2(f)	2	93	195	253	735	194	171

TABLE 2.1: Table showing the summary statistics for the different arrangements of a hypothetical mining area shown in figure 2.2.

Pebesma 1999, Best et al. 2001, Crawley & Harral 2001, Dorren & Heuvelink 2004, Jackson et al. 2006).

Finally, consideration of the support is important when the objective is to validate a model output against field data. For example, the remotely sensed data products are defined on pixel-sized support. However, field data gathered to validate such models may be gathered on substantially smaller supports. For example Wang et al. (2003) expresses concern that their field data (measured on a 1 m^2 support) used to validate their MODIS albedo product (defined on a 1 km $\times$ 1 km support) are defined on different supports. They use data from the Airborne Multiangular Thermal–Infrared Imaging System (AMTIS) (1.36 m spatial resolution) to provide an intermediate step to aggregate from 1 m to 1 km supports.

2.2.1.1 Aggregation

As discussed above, where exhaustive coverage of the variable is available aggregation is straightforward. This is demonstrated in figures 2.2(a) to 2.2(d), where the arithmetic mean is taken. However, this can be replaced with other quantities, such as the median or geometric mean, if required. However, the procedure may not be so straightforward for upscaling (Wen & Gomez-Hernandez 1996, Heuvelink & Pebesma 1999, Bierkens et al. 2000). Furthermore, the simplicity of the aggregation operation does not mean that there is no MAUP.

For most non-remotely sensed data sets coverage is not exhaustive. It is typical to have taken some type of spatial sample. This sample must then be used to predict the aggregated values. This situation is demonstrated in figure 2.2. Figure 2.2(a) shows the census, whereas figure 2.2(e) shows a sample. The objective is to predict the values of the 2 m $\times$ 2 m blocks, the true values of which are given in figure 2.2(b). There are various possibilities for predicting figure 2.2(b) from 2.2(e). The most straightforward is simply to use the one sample to predict the value for the larger block, as shown in figure 2.2(f), a procedure known as the "nearest neighbour" approach (Armstrong 1998, Burrough &

735	325	45	140	125	175	167	485
540	420	260	128	20	30	105	70
450	200	337	190	95	260	245	279
180	250	380	405	250	80	515	605
124	120	430	175	230	120	460	260
40	135	240	35	130	135	160	170
75	95	20	35	32	95	20	450
200	35	100	53	2	45	58	90

(a) 1 m × 1 m blocks

505	143	88	207
270	328	171	411
102	220	154	263
101	54	44	155

(b) 2m × 2 m blocks (mean of 1 m × 1 m blocks)

476	299	256	216	123	136	258	360
110	96	198	75	99	99	175	243

(c) 1 m × 4 m blocks (mean of 1 m × 1 m blocks)

311	238
337	56
294	220
304	363
212	268
113	149
26	149
97	49

(d) 4 m × 1 m blocks (mean of 1 m × 1 m blocks)

735							
		260		20		105	
	200		190				279
				250			
124							260
		240		130			
200			53	2		58	

(e) Sampled values for 1 m × 1 m blocks

735	260	20	105
200	190	250	279
124	240	130	260
200	53	2	58

(f) Predicted value for 2 m × 2 m blocks

FIGURE 2.2: Diagram showing the grades (unitless) in a mining area on 1 m × 1 m (2.2(a)) and 4 m^2 blocks (2.2(b), 2.2(c), 2.2(d)). The 4 m^2 blocks represent the mean value of different zonations of four 1 m^2 blocks. Figure 2.2(f) shows the predicted value of the 2 m × 2 m block based on sampling one of the values for 1 m × 1 m block (2.2(e). Adapted from (Armstrong 1998).

McDonnell 1998, Sadahiro 2000a, Martin & Atkinson 2001, Longley et al. 2005). However, comparison of figures 2.2(f) and 2.2(b) shows that this approach can be inaccurate. Furthermore, as shown in table 2.1, although the estimate of the regional mean approximates the true value, the standard deviation is over-estimated compared to the case where the true block values are used. Depending on the application, this inaccuracy may or may-not be unacceptable (Armstrong 1998). However, it begs the question of whether more accurate predictions can be obtained.

There are two broad classes of approaches for aggregating from support v to support V. The first are *design-based* methods and the second are techniques based on *interpolation*. Design-based approaches are based on taking a *random* sample. The block (V) value is predicted as a weighted average of the sample values within the block, where the weights are given by the probability of any individual sample being picked. If the probability of picking all samples is the same, this is simply the arithmetic average. This method requires that there is at least one sample per block (Heuvelink & Pebesma 1999) although, if the sample is too small, predictions may be inaccurate, as previously discussed. For larger sample sizes, the standard deviation can be calculated and the standard error of the estimate of the mean can also be calculated. Hence, the larger the sample size, the more accurate the block predictions will be. Design-based approaches have the advantage that they are conceptually simple and they are objective, in that they do not require any assumptions to be made about the spatial structure (Brus & de Gruijter 1997, Bierkens et al. 2000). However, as has been demonstrated, design-based approaches may be inaccurate for small samples. In many circumstances, it may also provide an inefficient sampling technique that does not exploit knowledge about the spatial structure. Furthermore, strictly it requires a random sampling technique which may not be feasible in practice (Bierkens et al. 2000).

The second class of approaches work by interpolation onto a surface and then taking an average of the predictions that lie within the block, V. As discussed in section 2.2 there are numerous methods available for interpolation. Geostatistical approaches are a common choice here (Heuvelink & Pebesma 1999, Dungan 1999, Bierkens et al. 2000). Such models are appropriate for non-randomly located data and utilise explicitly a model of spatial structure (Brus & de Gruijter 1997). As discussed later, the geostatistical approach of *block kriging* allows prediction onto a block with support V given input data on support v directly – that is the interpolation and aggregation steps are performed in one pass. However, if it is necessary to run the model on support v *before* aggregating, then it may be necessary to separate the interpolation and aggregation steps (Heuvelink & Pebesma 1999). Despite the advantages of geostatistical approaches it should be noted that they require an appropriately spaced large sample if the parameters of the spatial correlation structure are to be estimated accurately. Furthermore, there may be more subtle reasons why such approaches are not desirable, as discussed in section 2.2.2.6 and chapters 5 and 7.

Finally, Bierkens et al. (2000) present deterministic methods as a possibility for situa-

tions where geostatistical approaches are inappropriate. This involves using deterministic interpolation methods to predict onto a surface and then aggregating the predicted values that lie within V. For example, inverse distance weighting squared (IDW) could be used to predict onto a fine grid and an arithmetic average of the predicted points used to calculate the block predictor. Alternatively, Thiessen polygons could be drawn around the sample points and the weighted average calculated. The block value would then be given as a weighted average of all the points lying within the block, where the weight is given by the proportion of V that the Thiessen polygon occupies. This is analogous to the areal-weighting method often applied in GIS (Fisher & Langford 1995, Sadahiro 2000b). Deterministic methods are straightforward to implement and can be used under circumstances where there are insufficient or inappropriate data available to justify a design-based or geostatistically based approach. However, unlike these approaches, they do not provide any quantitative assessment of uncertainty.

2.2.2 Geostatistics

The term geostatistics has received some attention in previous sections. The objective of this section is to provide an outline of the geostatistical concepts that are most relevant for this thesis rather than to provide a detailed theoretical review. Geostatistics was originally developed in the mining industry (Matheron 1963, Journel & Huijbregts 1978, Armstrong 1998) and much of the literature that arose there is still widely used and referenced. However, geostatistics has also found wide application in the environmental sciences (Cressie 1993, Soares et al. 1996, Webster & Oliver 2001) including remote sensing (Curran & Atkinson 1998, Pardo-Igúzquiza et al. 2006) and geographical information science (Salvatori et al. 1999, Lark 2000, Lark & Cullis 2004, ESRI 2004). Much of the research into and practice of geostatistics has been conducted by applied scientists (e.g., Matheron 1963, Matheron 1965, Mátern 1986, Journel & Huijbregts 1978, Isaaks & Srivistava 1989, Goovaerts 1997, Armstrong 1998, Webster & Oliver 2001) rather than statisticians. However, statisticians have also had a role in the development of geostatistics (e.g., Cressie 1993, Harris & Johnson 1996, Diggle et al. 1998, Stein 1999, Cressie & Kornak 2003, Banerjee et al. 2004).

2.2.2.1 Random functions

Observed values at a set of points, $z(\mathbf{x})$, at locations $\mathbf{x}$ are considered as a realisation of a *random function* (RF) (also known as a random field or stochastic process), $Z(\mathbf{x})$. This is analogous to a random variable (RV). A realisation of a random function is termed the

regionalized variable (ReV). This extension from individual variables to fields is encapsulated in the *theory of the regionalized variable*, generally attributed to Georges Matheron (1963) (Oliver et al. 1989, Cressie 1993, Armstrong 1998, Webster & Oliver 2001). The random function is characterised by the joint distribution of its RVs where the spatial mean is termed the drift and is represented by the structural component (Journel & Huijbregts 1978, Armstrong 1998, Webster & Oliver 2001).

When data are realised from a RV, it is usual to obtain several data to make up a sample. This sample is then used to make inferences about the mean and variance (as well as higher order moments) of the population. In contrast, it is usual to obtain *one* ReV. Hence, in order to proceed and make inferences about the mean and covariance, an explicit assumption of *stationarity* of the RF is required (Webster & Oliver 2001). Strictly a RF is stationary if *all* its moments are invariant under translation. This requirement may not be met and cannot be verified from limited experimental data (Armstrong 1998). Hence, models have been developed on the basis of *second-order stationarity* which requires constancy of the covariance function and specifies that the covariance depends only on separation and not on absolute position.

Under second-order stationarity, the expected value is constant for all locations, $\mathbf{x}$:

$$E[Z(\mathbf{x})] = \mu \tag{2.48}$$

and the covariance between any two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is

$$Cov[Z(\mathbf{x}_i, \mathbf{x}_j) = E[(Z(\mathbf{x}_i) - \mu), (Z(\mathbf{x}_j) - \mu)] \tag{2.49}$$

Since, under second-order stationarity, the covariance is dependent on the separation $\mathbf{h} = \mathbf{x}_i - \mathbf{x}_j$, equation 2.49 becomes

$$\begin{aligned} Cov[Z(\mathbf{x}, Z(\mathbf{x}+\mathbf{h})] &= E[(Z(\mathbf{x}) - \mu), (Z(\mathbf{x}+\mathbf{h}) - \mu)] \\ &= E[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h}) - \mu^2] \\ &= C(\mathbf{h}) \end{aligned} \tag{2.50}$$

where $C(\mathbf{h})$ is the covariance at lag $\mathbf{h}$. The function $C(\cdot)$ is variously known as a *covariogram* or *covariance function* (Cressie 1993) or *autocovariance function* (Webster & Oliver 2001). The correlation (or autocorrelation) function is then given as

$$\rho(\mathbf{h}) = C(\mathbf{h})/C(\mathbf{0}) \tag{2.51}$$

where $C(\mathbf{0})$ is the covariance at lag 0, *i.e.*, the variance $\sigma^2 = E[Z(\mathbf{x}) - \mu]^2$

The condition of second-order stationarity may be weakened further to give the *intrinsic hypothesis*, which is attributed to Georges Matheron (Matheron 1963, Matheron 1965) (see also Armstrong 1998, Webster & Oliver 2001). This assumes stationarity of the *increment*,

such that:

$$E[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})] = 0 \tag{2.52}$$

and

$$\begin{aligned} Var[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})] &= E[\{Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})\}^2] \\ &= 2\gamma(\mathbf{h}) \end{aligned} \tag{2.53}$$

where $\gamma(\mathbf{h})$ is the *semi-variance* at lag $\mathbf{h}$ which, when plotted for different values of $\mathbf{h}$ gives the *semi-variogram* (now usually termed the variogram). The "semi" arises because the variogram is half the expected squared difference between two data separated by the lag $\mathbf{h}$.

Intrinsic stationarity applies under a wider range of situations than second-order stationarity. For example, it can be applied where the mean varies over the whole region but the difference is approximately zero over relatively short lags (Webster & Oliver 2001) and under conditions where a covariance function does not exist (Armstrong 1998). Note that second-order stationary RFs are a subset of intrinsically stationary RFs.

2.2.2.2 The (semi-)variogram

The (semi-)variogram is the fundamental tool for geostatistics (Cressie 1990, Oliver et al. 1989, Armstrong 1998, Curran & Atkinson 1998, Webster & Oliver 2001). It provides a tool for examining spatial structure and provides the basis for *kriging*. Kriging is used for spatial interpolation and is reviewed in section 2.2.2.3. The variogram can be used under more general circumstances than the covariance function (*i.e.*, intrinsic stationarity) (Cressie 1993, Armstrong 1998, Webster & Oliver 2001). Cressie also notes that, even under conditions of second-order stationarity, the variogram may be preferred since it can be shown to be less biased than the covariance function, particularly under conditions where any trend has not been fully modelled (Cressie 1993).

For a spatially referenced data set, the variogram shows how different the values are as the separation between them increases. The vector that defines the distance and direction of separation is termed the *lag* and denoted by the vector $\mathbf{h}$. The variogram was given in equation 2.53. If the RF is second-order stationary then the variogram is related to the covariance function, as follows:

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \tag{2.54}$$

A typical variogram is shown in figure 2.3 illustrating the three key parameters. These are the *range* (ϕ), the *sill* (also known as the total sill) ($\tau^2 + \sigma^2$) and the *nugget* (τ^2). The partial sill (σ^2) can be calculated if the sill and nugget are known. In theory, the stochastic

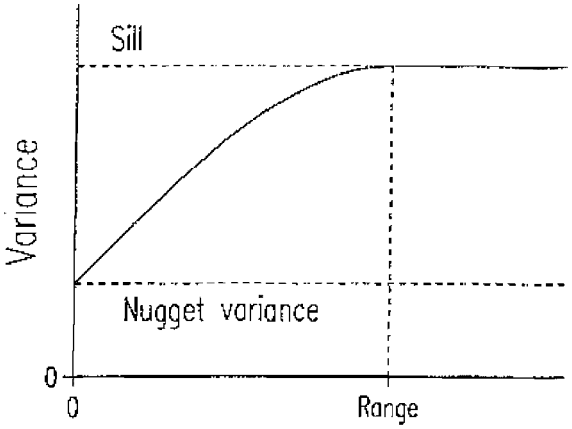


FIGURE 2.3: Example of a (spherical) variogram illustrating the three major parameters: the nugget, sill and the range. Reproduced from Webster & Oliver (2001), p. 56.

process giving rise to the variogram should be continuous everywhere, including where $\mathbf{h} = 0$. However, in practice, there is often a discontinuity as $\mathbf{h} \rightarrow 0$. This discontinuity is called the *nugget* and arises because of a combination of *(i)* structured variance over lags shorter than the minimum measurement interval and *(ii)* random error attributable to white noise or measurement error (Cressie 1993, Webster & Oliver 2001). It is usually not possible to distinguish between these two effects and it can be difficult to estimate the nugget accurately (Cressie 1993).

It is usual to distinguish between the *sample* and *theoretical* variogram. The term "experimental variogram" is widely used instead of "sample variogram". The term "sample variogram" is used in this thesis to avoid any confusion with scientific experimentation (Atkinson 1996). The sample variogram simply provides a way to plot the data and is defined as (Matheron 1963):

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} (Z(\mathbf{x}_i + \mathbf{h}_i) - Z(\mathbf{x}_i))^2 \tag{2.55}$$

where $Z(\mathbf{x}_i)$ is the value at location $\mathbf{x}_i$ and $N(\mathbf{h})$ is the number of pairs separated by lag, $\mathbf{h}$. When implementing this in practice it is common to specify a tolerance around the lag (the "bin"), especially where the data are not acquired at regular intervals. For two dimensional data, where the variogram may be calculated in more than one direction, it is also usual to specify an angular tolerance. Estimation of the sample variogram is sensitive to outliers, skew, sample size and sample spacing (Diggle et al. 1998, Webster & Oliver 2001, Diggle et al. 2003). Data may be examined for outliers and it may be possible to transform the data to ameliorate the effect of skew (Webster & Oliver 2001).

To gain an accurate estimate of the variogram for a two dimensional field Webster & Oliver (2001) recommend using *at least* 100 observations although they suggest that using approximately 150 or 250 observations will give a more accurate estimate for isotropic and non–isotropic data respectively. They propose that using a larger data set will, in general,

not offer a major advantage relative to the increased effort required. Regarding sample spacing, this must reflect the spatial structure of the surface in order to allow accurate estimation of the variogram at short lags, as well as accurate predictions based on kriging. In addition to the considerations of outliers, skew, sample size and sample spacing it should be noted that the choice of bin and angular tolerance can affect the appearance of the variogram (Diggle et al. 2003).

The sample variogram provides a subjective tool for examination of spatial structure. However, in order to quantify the nature of this structure it is necessary to fit a model. The estimated values of the nugget (τ^2) the partial sill (σ^2) and the range (ϕ) quantify the spatial variation. It is standard practice to fit a variogram model to the sample variogram (Journel & Huijbregts 1978, Armstrong 1998, Webster & Oliver 2001), although this approach is criticised by Diggle et al. (1998) and Diggle & Ribeiro Jr. (2002). The basis for that criticism is discussed in section 2.2.2.5.

Valid variogram models which may be fitted are sotimes referred to as *authorised* or *valid* models. Authorised models must be conditional negative semi–definite (CNSD) in order to ensure that the variances are not negative. Commonly used authorised models include the pure nugget model, the spherical model and the exponential model. The nugget model handles the case where there is no spatial structure. The spherical model is defined as:

$$\gamma(\mathbf{h}) = \begin{cases} \sigma^2 \left\{ \frac{3\mathbf{h}}{2\phi} - \frac{1}{2}\left(\frac{\mathbf{h}}{\phi}\right) \right\} & \text{for } \mathbf{h} \leq \phi, \\ \sigma^2 & \text{for } \mathbf{h} > \phi. \end{cases} \tag{2.56}$$

This gives rise to a bounded variogram, since the range is finite. The exponential model is defined as:

$$\gamma(\mathbf{h}) = \sigma^2 \left\{ 1 - \exp\left(-\frac{\mathbf{h}}{\phi} \right) \right\} \tag{2.57}$$

At first sight, the exponential model appears to give an unbounded variogram, since the semi-variance continues to rise until $\mathbf{h}$ approaches infinity. However, it is classified as a bounded variogram since, in the limit, it approaches $C(\mathbf{h})$ and the covariance function exists. Because the variogram continues to rise, it is usual to define an "effective" range, which is typically taken as the distance at which γ equals 95% of the sill variance and is approximately equal to 3ϕ. It is acceptable to combine (or nest) authorised models. Most commonly the nugget model is combined with another model.

It is common practice to fit an authorised variogram model to the sample variogram. This can be achieved by fitting the model by eye or by statistical fitting (Webster & Oliver 2001). Statistical fitting can be achieved using ordinary (OLS) or weighted least–squares (WLS) (Webster & Oliver 2001) and is implemented in commonly used geostatistical packages including `gstat` (Pebesma & Wesseling 1998, Pebesma 2003) and `geoR` (Ribeiro & Diggle 2001, Ribeiro Jr. et al. 2003). The weights are based on the number of data points that are used to estimate the variogram at each lag, although this may be further modified to give larger weight to the short lags. The fitted model should be checked

by visual comparison to the sample variogram. Furthermore, different variogram models may be compared using cross validation (see section 2.2.2.3) (Cressie 1993, Webster & Oliver 2001) and the Aikaike Information Criterion (AIC) (Akaike 1973) is often used to compare models (Webster & McBratney 1989, Webster & Oliver 2001). In common with other information criteria the AIC allows comparison between models with different numbers of parameters and, therefore, levels of complexity (Webster & McBratney 1989, Kuha 2004). The AIC has two terms, one which decreases as the prediction accuracy increases and the other that increases as number of parameters increases. Hence, the AIC compromises between accuracy and parsimony.

2.2.2.3 Kriging

Kriging refers to a suite of geostatistical techniques for interpolating between spatial sample points. Kriging is based on a continuous stochastic model that describes the form of the spatial variation and is applied to predict at a new location ($\mathbf{x}_0$) given data collected at neighbouring locations ($\mathbf{x}_i$). The spatial variation is described using by the variogram. There are several variants of kriging including ordinary kriging (OK) (where the mean is unknown), simple kriging (SK) (where the mean is known and assumed to be stationary over the region of interest) and universal kriging. Universal kriging will be developed in more detail in section 2.2.2.4. These techniques are all linear, although non-linear approaches also exist (Webster & Oliver 2001). In this section OK is described to provide a foundation that can be built on subsequently.

Kriging gives the best linear unbiased predictor (BLUP) (Cressie 1990). It is "best unbiased" because it minimises the error variance to give a mean residual of zero. It is "linear" because predictions are made using weighted linear combinations of the data. In the geostatistics literature, it is common to use "estimation" and "prediction" interchangeably (Webster & Oliver 2001), hence kriging is sometimes described as the best linear unbiased estimator (BLUE). However, in this thesis the statistical convention (Cressie 1990, Cressie 1993, Neter et al. 1996) of using estimation to describe parameter estimation and prediction to describe prediction of new values of a variable is applied.

For punctual kriging on the same support as the data, the predicted value is given as (Cressie 1990, Curran & Atkinson 1998, Webster & Oliver 2001):

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{N} \lambda_i z(\mathbf{x}_i) \tag{2.58}$$

such that $E[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = 0$. The λ_i are the kriging weights and sum to 1 and N is

the number of data observations in the region. The prediction variance is given as σ_E^2:

$$\begin{aligned}\sigma_E^2 &= E[\{\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)\}^2] \\ &= 2\sum_{i=1}^{N}\lambda_i\gamma(\mathbf{x}_i,\mathbf{x}_0) - \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j\gamma(\mathbf{x}_i,\mathbf{x}_j)\end{aligned} \tag{2.59}$$

where $\gamma(\mathbf{x}_i,\mathbf{x}_j)$ is the semi–variance between $\mathbf{x}_i$ and $\mathbf{x}_j$ and $\gamma(\mathbf{x}_i,\mathbf{x}_0)$ is the semi–variance between $\mathbf{x}_i$ and $\mathbf{x}_0$). This function is minimised by defining an auxiliary function containing a Lagrange multiplier and then setting the partial derivatives, with respect to the weights, to zero. This approach is described clearly by Webster & Oliver (2001) and elsewhere (Olea 1975). This yields a series of optimal weights (λ_is) that can be used to predict using equation 2.58. The kriging variance is then given as:

$$\sigma^2(\mathbf{x}_0) = \sum_{i=1}^{N}\lambda_i\gamma(\mathbf{x}_i,\mathbf{x}_0) + \psi(\mathbf{x}_0) \tag{2.60}$$

where ψ is the Lagrange multiplier.

This method can be extended to predict on a larger support, or block. Under this scenario, the predicted value for the block (B) is given as:

$$\hat{Z}(B) = \sum_{i=1}^{N}\lambda_i z(\mathbf{x}_i) \tag{2.61}$$

and the prediction variance is given as:

$$\begin{aligned}\sigma_E^2 &= E[\{\hat{Z}(B) - Z(B)\}^2] \\ &= 2\sum_{i=1}^{N}\lambda_i\bar{\gamma}(\mathbf{x}_i,B) - \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j\gamma(\mathbf{x}_i,\mathbf{x}_j) - \bar{\gamma}(B,B)\end{aligned} \tag{2.62}$$

where $\bar{\gamma}(\mathbf{x}_i,B)$ is the average semi–variance between the ith point and the block B (see figure 2.4 (a)) and is given by:

$$\bar{\gamma}(\mathbf{x}_i,B) = \frac{1}{|B|}\int_B \gamma(\mathbf{x}_i,\mathbf{x})d\mathbf{x} \tag{2.63}$$

and $\bar{\gamma}(B,B)$ is the average semi–variance between points within the block ($\mathbf{x}$ and $\mathbf{x}'$)(see figure 2.4 (b)) and is given by:

$$\bar{\gamma}(B,B) = \frac{1}{|B|^2}\int_B\int_B \gamma(\mathbf{x},\mathbf{x}')d\mathbf{x}d\mathbf{x}' \tag{2.64}$$

Equation 2.62 is minimised in the same way as equation 2.59. This allows calculation of the prediction (equation 2.61) and associated block kriging variance:

$$\sigma^2(B) = \sum_{i=1}^{N}\lambda_i\bar{\gamma}(\mathbf{x}_i,B) + \psi(B) + \bar{\gamma}(B,B) \tag{2.65}$$

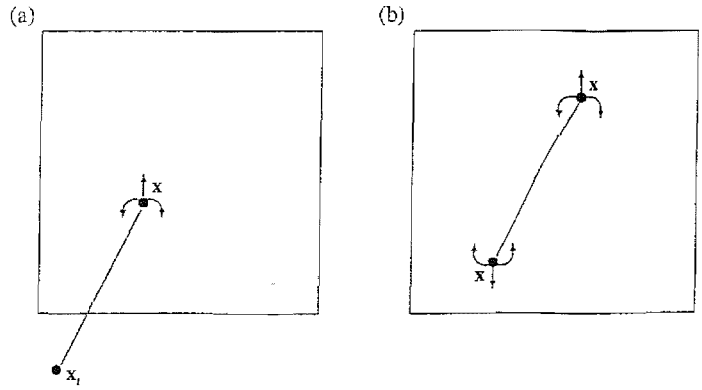


FIGURE 2.4: Integration of the variogram (a) between a sampling point and a block and (b) within the block. Reproduced from (Webster & Oliver 2001), p. 151.

For both punctual and block kriging, the equations can then be expressed in matrix form as:

$$\mathbf{A}\lambda = \mathbf{b} \tag{2.66}$$

where, $\mathbf{A}$ is a matrix containing each $\gamma(\mathbf{x}_i, \mathbf{x}_j)$, λ is a vector containing the kriging weights and a Lagrange multiplier and $\mathbf{b}$ is a vector containing $\gamma(\mathbf{x}_i, \mathbf{x}_0)$. The kriging weights are then derived by as:

$$\lambda = \mathbf{A}^{-1}\mathbf{b} \tag{2.67}$$

It is important to note that, since the kriging weights are based on the variogram, the kriging variance depends *only* on the variogram and the spatial separation of the data. It may also be noted that kriging is an *exact* interpolator and if prediction is performed at the location of a datum, that location is assigned a weight equal to one. A further point to note is that data that are close to the target location have greater weight than those that are far away. This highlights the fact that kriging is *local* (Webster & Oliver 2001). This has additional practical implications. First, only the locations close to the target carry significant weight which means that data that are far away may be disregarded without reducing the accuracy of prediction. This is of benefit computationally, since it reduces the size of the matrices that need to be inverted and reduces the risk of numerical instability (Pebesma & Wesseling 1998). If this approach is taken, care must be taken to ensure that data with relatively large weights are not omitted from the kriging system. Second, it means that the variogram needs to be estimated and modelled accurately at short lags, but inaccuracy at large lags can be tolerated, since it will not affect the kriging system. Hence, the notion of quasi–stationarity, whereby short–range structure is modelled but long–range structure is disregarded may be acceptable (Webster & Oliver 2001). Third, kriging within a local neighbourhood makes it possible to account for local variations in the mean. Hence, in many circumstances it can account for non-stationarity of the mean (Goovaerts 1997). The points raised in this paragraph are widely regarded as acceptable in the practice of geostatistics. However, they are not without controversy as is discussed in section 2.2.2.5.

The accuracy of prediction can be assessed through the process of cross validation. Under this approach a datum, $z(\mathbf{x}_i)$, is left out of the kriging system and its value is predicted at that location, $\hat{z}(\mathbf{x}_i)$. This procedure is repeated for each datum allowing comparison of the predicted and measured values. By extension, this approach also assesses the accuracy of the modelled variogram. Webster & Oliver (2001) give three heuristics. These are the Mean Error (ME) and Mean Square Error (MSE), introduced in equations 1.8 and 1.9 respectively, and the mean square deviation ratio ($MSDR$), given as

$$MSDR = \frac{1}{N}\sum_{i=1}^{N}\frac{\{z(\mathbf{x}_i) - \hat{z}(\mathbf{x}_i)\}^2}{\hat{\sigma}^2(\mathbf{x}_i)} \tag{2.68}$$

If the model is accurate then the MSE will equal the kriging variance and hence the $MSDR$ will tend to unity.

Finally, the conditional bias effect of kriging should be noted. Kriging *smoothes* and tends to over estimate at values of z that are low relative to the mean but under estimates at high values. Kriging gives the "best linear unbiased predictor" (BLUP) but does not maintain the variance that was present in the original data (Dungan 1999, Webster & Oliver 2001). Hence, a kriged surface is not a possible reality. Conditional simulation is an alternative to kriging that maintains the variance that was present in the original data. It is discussed in section 2.2.2.6.

2.2.2.4 Geostatistics and regression

Previous discussion has treated geostatistical techniques separately from regression modelling. However, geostatistical approaches can also be framed in terms of the linear regression model. The previous sections have also used the conventional geostatistical approach of writing the RF as $Z(\mathbf{x})$ and the ReV as $z(\mathbf{x})$. To maintain commonality with the regression model, I now switch notation and replace Z with Y, which was used to denote the response variable throughout section 2.1. The location $\mathbf{x}$ is replaced with $\mathbf{u}$, allowing x to be used to denote a predictor variable, as it was in section 2.1. This form of notation will be used for the remainder of the thesis.

Consider first the equation for the linear regression model (equation 2.1). This can be re-written to indicate the location of each data point:

$$y(\mathbf{u}_i) = \beta_0 + \sum_{k=1}^{p} x_i(\mathbf{u}_i)\beta_k + \varepsilon(\mathbf{u}_i) \tag{2.69}$$

In sections 2.2.2.1, 2.2.2.2 and 2.2.2.3 β_0 was the stationary mean and the second term on the RHS in equation 2.69 vanished. However, the second term does allow incorporation of further information. This information might take the form of a trend surface, where the xs are replaced with $\mathbf{u}$, or additional covariates. The third term represents the spatially

referenced residuals, which are likely to be spatially correlated. Utilisation of this approach for prediction may be referred to as *Universal Kriging (UK)* (Webster & Oliver 2001).

Having estimated the vector $\beta = (\beta_0, \beta_1, \ldots, \beta_p)$ and the variogram (or covariance function) of the residuals the value at a new location, $Y(\mathbf{u}_0)$, can be predicted as:

$$\hat{y}(\mathbf{u}_0) = \mathbf{X}(\mathbf{u}_0)\hat{\beta} + \boldsymbol{\Sigma}_0^T\boldsymbol{\Sigma}^{-1}(\hat{y}(\mathbf{u}_0) - \mathbf{X}\hat{\beta}) \tag{2.70}$$

where $x(\mathbf{u}_0)$ are the covariates at location $\mathbf{u}_0$, $\boldsymbol{\Sigma}$ is the covariance matrix for the residuals from equation 2.69 and, $\boldsymbol{\Sigma}_0 = (Cov[\hat{y}(\mathbf{u}_0), y(\mathbf{u}_1)], \ldots, Cov[\hat{y}(\mathbf{u}_0), y(\mathbf{u}_n)])^T$. Hence, the first term on the RHS is the prediction based solely on the regression model and the second term on the RHS modifies this value according to the values of the residuals at nearby data locations. The prediction error variance (kriging variance) is then given as:

$$\hat{\sigma}^2(\mathbf{u}_0) = \hat{\sigma}^2 - \boldsymbol{\Sigma}_0^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0 + (\mathbf{X}(\mathbf{u}_0) - \boldsymbol{\Sigma}_0^T\boldsymbol{\Sigma}^{-1}\mathbf{X})(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{X}(\mathbf{u}_0) - \boldsymbol{\Sigma}_0^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^T \tag{2.71}$$

which was presented in equation 2.60 in a different form. Equation 2.71 can be thought of as having three components. The first term is simply the error variance. The second term has a negative sign because, having measurements at neighbouring points *reduces* uncertainty in the prediction. However, the third term has a positive sign because uncertainty in the estimation of β acts to *increase* the uncertainty of prediction. If the measured values are very far from $\mathbf{u}_0$ (beyond the variogram range), then the second term in equations 2.70 and 2.71 will approach zero. If β is known *a priori* (as in simple kriging) and do not need to be estimated, the third term in equation 2.71 will be zero.

The prediction equations presented in equations 2.70 and 2.71 require that both β and variogram (or covariance function) have been estimated. However, the estimation of these parameters leaves a "chicken and egg" situation. This is because β is estimated as:

$$\hat{\beta}_{gls} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} \tag{2.72}$$

which requires $\boldsymbol{\Sigma}$. However, the variogram is required in order to calculate the elements of $\boldsymbol{\Sigma}$ and the residuals from equation 2.69 are required to estimate the variogram. One approach to dealing with this is to begin by estimating β using OLS (which does not require $\boldsymbol{\Sigma}$) and to use the OLS residuals to estimate the variogram and therefore $\boldsymbol{\Sigma}$. This can then be input into equation 2.72 and used to calculate the GLS estimate of β. This can be iterated until convergence. Often the OLS estimators of β are used in place of the GLS estimators since, although they are sub-optimal, they are often not substantially different (Pebesma & Wesseling 1998, Pebesma 2003, Hengl et al. 2004, Lark & Webster 2006). Hence `gstat` by default works with the OLS estimators, although the user can calculate GLS estimators if required. However, the matrix $C(\hat{\beta})$ (see also equation 2.38) will differ for the OLS and GLS estimators.

The implementation described above involves calculating either a global mean (i.e., the second term in equation 2.69 does not exist) or a global model for the trend – based either

on location or on a covariate. This approach is advocated by some researchers (Pebesma & Wesseling 1998, Diggle et al. 1998, Diggle & Ribeiro Jr. 2002, Diggle et al. 2003, Pebesma 2003, Lark & Cullis 2004, Hengl et al. 2004, Lark & Webster 2006) and is implemented in the widely used `gstat` and `geoR` software packages. It is sometimes referred to as *Regression Kriging (RK)* (Knotters et al. 1995, Odeh et al. 1995, Hengl et al. 2003, Hengl et al. 2004, Lark & Webster 2006). However, this is not the only approach available. It is common to distinguish between a trend on location (commonly known simply as a trend) and a covariate (commonly known as an "external drift") (Journel & Huijbregts 1978, Goovaerts 1997, Goovaerts 1999). This leads to kriging with a trend (KT) and kriging with an external drift (KED) respectively. For both KT and KED, Goovaerts (Goovaerts 1997, Goovaerts 1999) advocates estimating the variogram from data, $\mathbf{y}(\mathbf{u})$ that are not strongly affected by the trend or the drift. This variogram (or the associated covariance function) is then used in a system of kriging equations (solved using the Lagrange multiplier approach). Goovaerts (1997) recommends defining a local neighbourhood for kriging. Hence the kriging equations for KT and KED lead to local, rather than global, estimates of β. Depending on the situation, this may be more appropriate than a global estimate. Indeed, KED or KT will give the same predictions as RK if a global neighbourhood is used. Finally, it may be noted that implementing RK requires that β is calculated explicitly whereas, for KT or KED, they are calculated as part of the kriging system (although they can be obtained if required).

It should be noted that a trend can also be accounted for using OK – whereby kriging is implemented in a local neighbourhood using a variogram that has been estimated for $\mathbf{Y}(\mathbf{u})$ rather than $\varepsilon(\mathbf{u})$. This is because prediction is *local* and, hence, can handle longer range non-stationarity. Goovaerts (1997) and Webster & Oliver (2001) note that this may be sufficient to account for a trend and shows greater parsimony than resorting to KT. However, Goovaerts (1997) notes that KT may be preferred if there is a trend *within* the local neighbourhood.

Finally, Goovaerts (1997) (see also: Goovaerts 1999) presents simple kriging with varying local means (SKlm) as an alternative to KED. SKlm can be implemented for categorical or continuous covariates. In the former case, the mean varies locally according to the class. In the second case, the mean is given as a function of the covariates. This function could be a physical model or, more usually, a regression model of the type presented in equation 2.69. Where a regression model is adopted SKlm is analogous to RK where the OLS (rather than GLS) estimates of β are used.

2.2.2.5 Model–based geostatistics

The discussion of geostatistics has, so far, focused on "classical" or "conventional" geostatistics. which has developed largely in geology and environmental sciences. Model–based geostatistics (MBG) represents an alternative, but related, approach to geostatistics. The term "model–based" emphasises that the approach is based on explicitly de-

clared stochastic models and associated formal methods of statistical inference (Diggle et al. 1998). There is considerable cross–over with the classical approach, although there are epistemological differences and practical differences. The approach is outlined below and this incorporates a discussion of the major differences in principle and practice. This discussion is important, given that this thesis focuses on uncertainty. This section draws heavily on the work of Diggle and co-workers (Diggle et al. 1998, Ribeiro & Diggle 2001, Christensen & Ribeiro 2002, Diggle & Ribeiro Jr. 2002, Diggle et al. 2003, Ribeiro Jr. et al. 2003). Their material is clearly presented and they have also undertaken pioneering work concerning Bayesian approaches and general linear models. Several other authors have undertaken research in a similar framework. For example, maximum likelihood (e.g., Cressie 1993, Harris & Johnson 1996, Gotway & Stroup 1997, Lark 2000, Ver Hoef et al. 2001, Lark & Cullis 2004) and Bayesian (e.g., Omre 1987, Banerjee et al. 2004) approaches are presented elsewhere.

The MBG approach is described in detail by Diggle et al. (1998) and Diggle et al. (2003). MBG can be implemented using the `geoR` package for the `R` statistical programming environment (Ribeiro & Diggle 2001, Ribeiro Jr. et al. 2003). The underlying model is given by:

$$Y(\mathbf{u}_i) = S(\mathbf{u}_i) + Z_i \tag{2.73}$$

where the *sampling design* is given by $\{\mathbf{u}_i : 1, 2, \ldots, N\}$ and $\{Y(\mathbf{u}) : \mathbf{u} \in R\}$ is the *measurement process.* An underlying stochastic process, termed the *signal process* $\{S(\mathbf{u}) : \mathbf{u} \in R\}$ is assumed, where $Y(\mathbf{u}_i)$ is often a noisy version of $S(\mathbf{u}_i)$. Hence Z_i describes the uncorrelated error and is not dependent on $\mathbf{u}$. It will be shown that this is the same as equation 2.69.

Prediction is often the key objective. The random variable T is denoted as the target for prediction given $\mathbf{y}$ which is a realisation of $\mathbf{Y}$. The minimum mean square error predictor for T is given as:

$$\hat{T} = E[T|\mathbf{y}] \tag{2.74}$$

which minimises:

$$MSE(\hat{T}) = E[(T - \hat{T})^2] = E_Y[Var(T|\mathbf{y})] \tag{2.75}$$

The above description makes no distributional assumption. However, it is usual to model $S(\cdot)$ is as stationary Gaussian process with:

$$\begin{aligned} E[S(\mathbf{u})] &= \mu \\ Var[S(\mathbf{u})] &= \sigma^2 \\ \rho(\mathbf{h}) &= Corr[S(\mathbf{u}), S(\mathbf{u}')] \\ Cov[S(\mathbf{u}), S(\mathbf{u}')] &= \sigma^2\rho(\mathbf{h}) \end{aligned} \tag{2.76}$$

Furthermore, the conditional distribution of $Y(\mathbf{u}_i)$ given $S(\cdot)$ is normally distributed with

mean $S(\mathbf{u}_i)$ and variance τ^2, i.e.:

$$Y(\mathbf{u}_i)|S(\cdot) \sim N(S(\mathbf{u}_i), \tau^2) \tag{2.77}$$

which can also be expressed as:

$$Y(\mathbf{u}_i) = S(\mathbf{u}_i) + \varepsilon \tag{2.78}$$

where $\varepsilon = N(0, \tau^2)$. Hence, τ^2 can be thought of as the nugget variance. The joint distribution of $\mathbf{Y}$ is multivariate normal:

$$\mathbf{Y} \sim MVN(\mu\mathbf{1}, \sigma^2\mathbf{R} + \tau^2\mathbf{I}) \tag{2.79}$$

where $\mathbf{R}$ is an $N \times N$ matrix where the (i, j)th element is $\rho(||\mathbf{u}_i - \mathbf{u}_j||)$. In equation 2.78, $\sigma^2\mathbf{R} + \tau^2\mathbf{I}$ is equivalent to the covariance matrix, $\mathbf{\Sigma}$, presented in the previous section and $C(\mathbf{0}) = \sigma^2 + \tau^2$. Note also that μ need not be constant and can be modelled as $\mu = \beta\mathbf{X}$, where $\mathbf{X}$ may be location (i.e., modelling a trend) or a covariate, as discussed in section 2.2.2.4.

The correlation function describes the smoothness of the process, $S(\cdot)$. Particular attention is given to the Mátern (1986) class of models (Diggle et al. 1998, Diggle et al. 2003):

$$\rho(\mathbf{h}; \phi, \kappa) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(\mathbf{h}/\phi)^{\kappa}K_{\kappa}(\mathbf{h}/\phi) \tag{2.80}$$

which is valid for $\phi > 0$ and $\kappa > 0$ and where $K_{\kappa}(\cdot)$ denotes a Bessel function of order κ. The exponential model (given in equation 2.57) is a special case where $\kappa = 0.5$. This model is preferred (Diggle et al. 1998, Diggle et al. 2003) because of its flexibility and because it is mean square continuous or mean square differentiable (depending on κ), making the function amenable to computation when likelihood-based methods are used (Diggle & Ribeiro Jr. 2002). However, other models, such as the spherical model, may be chosen.

If the correlation model and its parameters are known, prediction may be performed at a new location, $\mathbf{u}_0$:

$$\hat{T}_0 = \hat{S}(\mathbf{u}_0) = \mu_0 + \sigma^2\mathbf{r}^T(\tau^2\mathbf{I} + \sigma^2\mathbf{R})^{-1}(\mathbf{y} - \mu) \tag{2.81}$$

with prediction variance:

$$Var(T_0|\mathbf{y}) = \sigma^2 - \sigma^2\mathbf{r}^T(\tau^2\mathbf{I} + \sigma^2\mathbf{R})^{-1}\sigma^2\mathbf{r} \tag{2.82}$$

where $\mathbf{r}$ is vector with elements $\mathbf{r}_i = \rho(||\mathbf{u}_i - \mathbf{u}_0||)$. Note the similarity between this equation and equation 2.71. However, this equation assumes that the mean, μ, is known hence the third term in equation 2.71 tends to zero. As before, this system can be extended such that $\mu = \mathbf{X}\beta$.

Diggle et al. (2003) are critical of the standard approach of estimating the model param-

eters (τ^2, σ^2 and ϕ) by fitting a model to the sample variogram. First, they emphasise that the sample variogram is unstable, particularly at large lags, and is sensitive to the choice of bins. Second, they describe the conventional approach of parameter estimation as "ad hoc" and favour methods based on the likelihood, since they allow recourse to established theory and practice in likelihood-based estimation. Notably, for large samples, likelihood–based methods are optimal, in the sense that they are minimum variance unbiased. However, they do note that the optimality properties may be difficult to interpret for spatially correlated data (Diggle & Ribeiro Jr. 2002) and may lack robustness (Diggle et al. 2003). Furthermore, they can be computationally demanding. Noting equation 2.79, the log-likelihood is:

$$\begin{aligned} l(\mathbf{y};\beta,\tau^2,\sigma^2,\phi,\kappa) &\propto -\frac{1}{2}\left\{\log|\sigma^2\mathbf{R}+\tau^2\mathbf{I}| + (\mathbf{y}-\beta\mathbf{X})^T(\sigma^2\mathbf{R}+\tau^2\mathbf{I})^{-1}(\mathbf{y}-\beta\mathbf{X})\right\} \\ &= \frac{n}{2}\log\sigma^2 + \frac{n}{2}\log|\mathbf{V}| + \frac{1}{2\sigma^2}(\mathbf{y}-\beta\mathbf{X})^T\mathbf{V}^{-1}(\mathbf{y}-\beta\mathbf{X}) \qquad (2.83) \end{aligned}$$

where $\mathbf{V} = (\sigma^2\mathbf{R} + \tau^2\mathbf{I})/\sigma^2$. The model can also be extended to include additional parameters to account for anisotropy. Maximisation of this likelihood proceeds as follows

1. Select initial values for τ^2, σ^2 and θ.

2. Re-parameterise to $\nu^2 = \tau^2/\sigma^2$ and set $\mathbf{V} = \mathbf{R}+\nu^2\mathbf{I}$. This re-parameterisation is for computational reason and yields a modified correlation matrix. A similar approach is adopted elsewhere (Cook & Pocock 1983, Harris & Johnson 1996, Lark 2000). Lark & Cullis (2004) adopt a slightly different re-parameterisation such that $\nu = \tau^2/(\tau^2+\sigma^2)$.

3. The parameters can be thought of as comprising two groups. First, $\beta = (\beta_0, \ldots, \beta_p)^T$ defines the trend (or the mean). Second, $\theta = (\sigma^2, \nu^2, \phi)^T$ defines the covariance function.

4. Calculate $\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$ as in equation 2.72.

5. Calculate $\hat{\sigma}^2$, the ML estimator of σ as:

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y}-\hat{\beta}\mathbf{X})^T\mathbf{V}^{-1}(\mathbf{y}-\hat{\beta}\mathbf{X}) \qquad (2.84)$$

6. Substitute $\hat{\beta}$ and $\hat{\sigma}^2$ into the log-likelihood. However, since this is now maximised with respect to $\hat{\beta}$ and $\hat{\sigma}^2$, the third term on the RHS of equation 2.83 depends only on n. Hence, equation 2.83 reduces to:

$$l^*(\nu^2,\phi,\kappa) \propto -\frac{1}{2}(n\log\sigma^2 + \log|\mathbf{V}|) \qquad (2.85)$$

7. Record and store the value of $l^*(\nu^2,\phi,\kappa)$.

8. Return to Step 4 until equation 2.85 is maximised.

This approach is implemented by a range of authors (e.g., Cook & Pocock 1983, Mardia & Marshall 1984, Harris & Johnson 1996, Lark 2000). It is possible to choose between models, such as those with different numbers of β parameters and those with different correlation functions, by comparing the log-likelihoods as well as other measures such as the AIC and BIC.

Maximum likelihood approaches are known to lead to biased estimates of the variance parameters, σ^2, as discussed in section 2.1.3. This bias arises because these parameters are dependent on β (Cressie 1993, Diggle et al. 2003, Lark & Cullis 2004). Note that a correction can be applied to the ML estimator of σ^2 such that (Harris & Johnson 1996, Ver Hoef et al. 2001):

$$\tilde{\sigma}^2 = \frac{n\hat{\sigma}^2}{n-p} \tag{2.86}$$

where p is the length of the vector, β. However, in simulation experiments Harris & Johnson (1996) still find $\tilde{\sigma}^2$ to be biased.

Restricted maximum likelihood (REML) (also known as residual maximum likelihood) has been proposed as a technique for reducing the effect of bias (Cressie 1993, Gilmour et al. 1995, Stuart et al. 1999, Diggle et al. 2003, Lark & Cullis 2004). This works by defining a new likelihood which is conditional on the parameters in β. Following some algebraic manipulation the REML log-likelihood is given as (Stuart et al. 1999):

$$l(\theta|\hat{\beta},\beta) \quad \propto \quad -\frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\log|\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{Y}^T\mathbf{\Sigma}^{-1}(\mathbf{I}-\mathbf{Q})\mathbf{y} \tag{2.87}$$

where

$$Q \equiv \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}$$

Use of REML over ML is advocated by several authors (Cressie 1993, Smyth & Verbyla 1996, Lark & Cullis 2004). However, Diggle et al. (2003) note that this applies particularly to small samples and that, in their experience it is more sensitive than ML to mis-specification of the trend. This is given further consideration in the analysis chapters.

Following estimation of the parameters ($\theta = \tau^2$, σ^2 and ϕ), prediction may be performed using kriging, as described in section 2.2.2.3. This "plug-in-prediction" approach mirrors that used in conventional geostatistics and in regression modelling more generally. The expectation and kriging variance do quantify uncertainty, but this is based on fixed estimators for θ. Hence, a criticism that can be levelled at "plug-in-prediction" is that it does not account for uncertainty in the estimation of θ. However, this criticism can be interpreted more generally as a criticism of classical statistics, which treats parameters as fixed unknown values. In comparison, Bayesian statistics allows for uncertainty in the parameter estimates (Gelman et al. 1995, Gilks et al. 1996). Diggle et al. (2003) note that failure to account for parameter uncertainty tends to give accurate point predictions, but tends to underestimate the prediction variance and may give inaccurate predictions of other properties, such as thresholds. Hence, they advocate use of a Bayesian paradigm.

Moving to the Bayesian inferential paradigm is advantageous, since it provides a natural means of formally incorporating prior information about the model parameters and because uncertainty in the model is incorporated explicitly into prediction. Consider the parameter vector $\theta = (\beta, \tau^2, \sigma^2, \phi, \kappa)$. Under the Bayesian framework (e.g., Gelman et al. 1995, DeGroot & Schervish 2002) the posterior distribution is given as:

$$\begin{aligned} \pi(\theta|\mathbf{y}) &\propto p(\mathbf{y}|\theta)\pi(\theta) \\ \textit{Posterior} &\propto \textit{Likelihood} \times \textit{Prior} \end{aligned} \tag{2.88}$$

where $\pi(\theta)$ is the prior distribution of θ and $p(\mathbf{y}|\theta)$ is the likelihood, which might also be expressed as $L(\mathbf{y};\theta)$. The joint posterior predictor for T is then given as:

$$p(\mathbf{T}|\mathbf{y}) = \int_{\Theta} p(\mathbf{T}|\mathbf{y},\theta)\pi(\theta|\mathbf{y})d\theta \tag{2.89}$$

The distribution $p(\mathbf{T}|\mathbf{y})$ may have an analytical expression or may be found by simulation, allowing calculation of measures such as the mean, variance or 95% intervals. The choice of priors is recognised as a delicate issue in Bayesian inference. It is mathematically and computationally convenient to chose a conjugate prior, since this leads to a posterior in the same family of distributions (DeGroot & Schervish 2002). In the MBG case, Diggle & Ribeiro Jr. (2002) have developed a system that requires choosing from a limited set of conjugate priors for β and σ^2, but uses a uniform, discrete joint prior for (ϕ, τ^2). The implementation is described as follows, for the case where $\tau^2 = 0$, and ϕ and κ are fixed. The posterior density of $\theta = (\beta, \sigma^2)^T$ (β is the vector of regression coefficients) is given as:

$$\begin{aligned} \pi(\beta, \sigma^2|\mathbf{y}, \phi) &\propto p(\mathbf{y}|\beta, \sigma^2, \phi)\pi(\beta, \sigma^2|\phi) \\ \textit{Posterior} &\propto \textit{Likelihood} \times \textit{Prior} \end{aligned} \tag{2.90}$$

where the conjugate prior for $\pi(\beta, \sigma^2|\phi)$ is the Gaussian–Scaled–Inverse–χ^2 ($N\chi^2_{ScI}$), which specifies priors for β and σ^2 as:

$$\begin{aligned} \pi(\beta|\sigma^2, \phi) &\sim N(\mathbf{m}_b, \sigma^2\mathbf{V}_b) \\ \pi(\sigma^2|\phi) &\sim \chi^2_{ScI}(n_{\sigma^2}, S^2_{\sigma^2}) \end{aligned} \tag{2.91}$$

where χ^2_{ScI} is the Scaled–Inverse–χ^2 distribution. These are then combined in a hierarchical framework to yield:

$$\pi(\beta, \sigma^2|\phi) \sim N\chi^2_{ScI}(\mathbf{m}_b, \mathbf{V}_b, n_{\sigma^2}, S^2_{\sigma^2}) \tag{2.92}$$

This is conjugate when combined with the Gaussian likelihood function, giving:

$$\pi(\beta, \sigma^2|\mathbf{y}, \phi) \sim N\chi^2_{ScI}(\tilde{\beta}, \mathbf{V}_{\tilde{\beta}}, n_{\sigma^2} + n, S^2) \tag{2.93}$$

where $\tilde{\beta}$, $\mathbf{V}_{\tilde{\beta}}$ and S^2 have analytical solutions (see Diggle et al. (2003) for details) and n

is the number of data points. The predictive distribution is then given as:

$$p(\mathbf{T}|\mathbf{y},\phi) = \int\int f(\mathbf{T}|\mathbf{y},\beta,\sigma^2,\phi)\pi(\beta,\sigma^2|\mathbf{y},\phi)d\beta d\sigma^2 \tag{2.94}$$

which gives a multivariate t-distribution:

$$f(\mathbf{T}|\mathbf{y},\phi) \sim t_{n_{\sigma^2}+n}(\mu_{\mathbf{T}}, S^2\mathbf{V}_{\mathbf{T}}) \tag{2.95}$$

A possible non–informative prior is $\pi(\beta,\sigma^2|\phi) = 1/\sigma^2$, in which case $\mathbf{V}_b = 0$, $n_{\sigma^2} = 0$ and $n_{\sigma^2} + n$ is replaced by $n - p$.

This can be generalised to allow for uncertainty in ϕ:

$$\pi(\beta,\sigma^2,\phi|\mathbf{y}) = \pi(\beta|\sigma^2,\phi)p(\phi|\mathbf{y}) \tag{2.96}$$

where:

$$p(\phi|\mathbf{y}) \propto \pi(\phi)|\mathbf{V}_{\tilde{\beta}}|^{1/2}|\mathbf{R}|^{-1/2}(S^2)^{\frac{n+n_{\sigma^2}}{2}} \tag{2.97}$$

where $\mathbf{V}_{\tilde{\beta}}$ and S^2 are found from equation 2.94. $\pi(\phi)$ is discretised on a user specified range, and this is used to compute $p(\phi|\mathbf{y})$. This is then sampled from and the value of ϕ is attached to equation 2.90. Repeated sampling builds up the sample, $\pi(\beta,\sigma^2,\phi|\mathbf{y})$. The posterior predictive distribution is then given as:

$$\begin{aligned} f(\mathbf{T}|\mathbf{y}) &= \int\int\int f(\mathbf{T}|\mathbf{y},\beta,\sigma^2,\phi)\pi(\beta,\sigma^2,\phi|\mathbf{y})d\beta d\sigma^2 d\phi \\ &= \int\int\int f(\mathbf{T}|\mathbf{y},\beta,\sigma^2,\phi)\pi(\beta,\sigma^2|\mathbf{y},\phi)d\beta d\sigma^2 f(\phi|\mathbf{y})d\phi \\ &= \int f(\mathbf{T}|\mathbf{y},\phi)f(\phi|\mathbf{y})d\phi \end{aligned} \tag{2.98}$$

which combines the uncertainty in ϕ with equation 2.94.

The above scheme can be generalised for the situation where τ^2 also needs to be estimated. This is achieved by applying the discretisation to a joint prior $\pi(\phi,\nu^2)$, where $\nu^2 = \tau^2/\sigma^2$. This results in the same procedure as that given above, but the computational load is increased.

The above discussion has drawn out both similarities and differences with classical geostatistics. These differences are summarised below.

1. The MBG framework formally and explicitly states the theoretical and epistemological framework for estimation and prediction in a geostatistical setting.

2. The MBG framework rejects the conventional approach of fitting a model to the sample variogram in favour of likelihood based methods (including Bayesian methods). Strong mathematical and statistical reasons are provided for this. However, likelihood based methods are not without their problems. First, they may lack robustness to outliers. Second, the spatial correlation means that the effective sam-

ple size is smaller than the actual sample size, and this may affect the optimality properties.

3. There is no formal requirement for kriging to honour the data – a point that contradicts conventional geostatistical implementation described previously . Note that this depends on whether τ^2 is interpreted as measurement error or as micro-scale variation (Cressie 1993, Diggle & Ribeiro Jr. 2002, Diggle & Ribeiro Jr. 2006). If τ^2 is taken to represent micro–scale variation, then predicted points will honour the data.

4. Under the model–based approach, the practice of restricting the neighbourhood for prediction is not regarded as being theoretically desirable, although it may have computational benefits.

5. The Bayesian approach can be considered to have both practical and epistemological benefits. A pure-bred Bayesian might argue that this approach is *de facto* superior to the classical ML or REML approach. The benefits of the Bayesian approach can also be seen from a practical perspective, since they allow for construction of hierarchical models that incorporate uncertainty in parameter estimation into prediction (kriging). Classical approaches *do* allow the analyst to account for uncertainty through calculation of the variance as well as confidence and prediction intervals. However, these approaches can break down in the spatial context since they are sensitive to biased estimation of ϕ (Harris & Johnson 1996) leading to inaccurate estimates of the prediction variance (Diggle et al. 2003). By contrast the Bayesian approach does allow uncertainty in estimation of all the model parameters. However, it should be noted that this may not always be of benefit. As discussed in section 2.2.2.3 the (conventional) kriging variance only depends on the separation of the data points and the variogram model. If the objective is to isolate the effect of the sampling strategy for a given variogram model, then plug–in prediction would be acceptable.

It should be recognised that the practice of fitting a model to the sample variogram and of restricting the neighbourhood for prediction allows the "conventional" geostatistician flexibility, which can be used effectively by a person who understands the data and the methodology. This is because they can choose to restrict the neighbourhood for prediction and maximise the precision of the model fit at short lags, thus ignoring long range structure. This notion of "quasi–stationarity" is acceptable in many situations (Webster & Oliver 2001). Such *ad–hoc* methods may be justified where they improve the accuracy of prediction (Journel & Huijbregts 1978, Goovaerts 1997, Goovaerts 1999, Webster & Oliver 2001). The model–based approach denies this flexibility to the practitioner, because it does not fit in with the formal statement of the model. Hence, they will have to find other ways to reach the desired outcome. In the Bayesian context, the choice of priors affects both parameter estimation and prediction. In the procedure described above, this is particularly noticeable for τ^2 and ϕ. The choice of the range over which $\pi(\nu^2, \phi)$ is discretised can have a large effect on the estimation of β, τ^2, σ^2 and ϕ. The

user may have information on measurement error that can be used to inform the choice of the range for ν^2. However, Diggle & Ribeiro Jr. (2002) and Diggle et al. (2003) provide no clear guidance as to how to determine the range for ϕ. It should also be noted that there is a computational trade–off between the range over which the prior is discretised and the denseness of the discretisation. This is because increasing the number of points in the discrete sample has an associated computational cost. The implication here is that the user can spuriously affect the parameter estimates through choosing an inappropriate prior or because the density of the discretisation is insufficient. The correct choice of prior may be difficult to justify and the results difficult to visualise. The latter arises because a variogram that is estimated by likelihood methods may, quite correctly, not resemble the sample variogram (Diggle et al. 2003, Lark & Cullis 2004).

In summary, MBG does make a valuable contribution to geostatistics. Notably, it formalises the use of the model, allows incorporation of the Bayesian paradigm and allows uncertainty in parameter estimation to be incorporated into prediction uncertainty. However, it also brings disadvantages. First, there is no clear guidance for selection of priors, particularly for τ^2 and ϕ. Choices made here affect both estimation and prediction. Second, it denies the conventional geostatistician his or her *ad hoc* tools without providing a clear alternative. The model–based geostatistician may regard this as a good thing, since they deviate from the model. Subject matter knowledge may be incorporated through the choice of variogram model, choice of prior and choice of any "add ons" (such as models to deal with anisotropy or non–stationarity), although mechanisms for doing this are not always clear. It should also be noted that the conventional geostatistician also places stringent requirements on the application of these ad hoc techniques. Notably, they must be shown to "work" in practice including in scenarios where economic stakes are high (Journel & Huijbregts 1978, Armstrong 1998). Finally the computational burden of implementing MBG may be greater than for conventional geostatistics.

2.2.2.6 Stochastic simulation

As discussed in section 2.2.2.3 kriging yields the best linear unbiased predictor (BLUP). Hence, a surface that is predicted using kriging will be *smooth* relative to the real surface (Journel 1996b, Goovaerts 1997, Atkinson 1999b, Dungan 1999). This can be thought of by analogy to a linear regression line. A prediction made using the regression model will lie on the "best fit" regression line, whereas the data used in the parameter estimation stage will be scattered about that regression line.

A consequence of this smoothing is that relatively large values will tend to be under-predicted and small values over-predicted. Furthermore, this smoothing is not uniform since the level of smoothing will be lower close to data points than it will be far from data points (Goovaerts 1997). A kriged surface will tend to have different second-order statistical properties from the real surface and the original data. In particular, the histogram of predicted values will show a narrower range of values, hence the variance will be lower.

These differences in statistical properties mean that, although the kriged surface is an optimal predictor, it does not yield a surface that is a "possible reality". Even when accompanied by the associated map of the kriging variance, the kriged surface may not be suitable in many situations. These differences are classified by Dungan (1999), as follows:

1. It may be necessary to make statements about a *group* of pixels, rather than about individual pixel values or locations. For example, one might want to estimate the probability of a group of pixels exceeding a particular value (Deutsch & Journel 1998).

2. It may be necessary to combine the predicted surface with data in another GIS layer. For example, the objective may be necessary to combine layers in an empirical statistical model. If the layers are from kriged maps then the smoothing will influence the correlation between the layers. Furthermore, where the surface is to be used as an input into a physically based model (such as a hydrological model), using a kriged surface may give an unrealistic assessment of the range of possible model outputs (Heuvelink 1998, Atkinson & Tate 2000).

3. A more extensive examination of uncertainty than that provided by examination of the kriging variance may be required.

Stochastic simulation (also known as stochastic imaging (Journel 1996b)) aims to produce a surface that, whilst it is sub-optimal, does represent a possible reality. This is achieved by generating a surface that maintains the properties of the variogram (or covariance function) used to create it. It is usual to generate multiple realisations of the surface to explore different scenarios in the above contexts.

Conditional simulation is undertaken by conditioning on data, whereas *unconditional simulation* does not require conditioning on data (Heuvelink 1998, Dungan 1999). However, both forms of simulation should aim to maintain the variogram properties.

Dungan (1999) provides three classes of conditional simulation algorithm that are broadly reflected by other authors (Journel 1996b, Heuvelink 1998, Atkinson 1999b). These are

1. *Strategies based on adding the missing spatial variability "back in".* This approach is based on adding spatially correlated variability, generated by unconditional simulation, back to the kriged estimate. Algorithms include the turning band method (Journel 1973) and approaches based on the Cholesky (or LU) decomposition (Scheuer & Stoller 1962, Johnson 1987, Davis 1987b, Davis 1987a). The turning bands method has the advantage of computational efficiency, but offers little flexibility and the simulations may contain artifacts (Dungan 1999). The Cholesky decomposition approach is widely used to generate correlated variables (Carsel & Parrish 1988, Ferson et al. 2004) but can be problematic in the spatial context owing to the potentially very large matrices.

2. *Sequential approaches.* This approach is based on Bayesian principles whereby a realisation is drawn from a conditional cumulative density function (ccdf) at a randomly determined location, where the conditioning is on the observed data. This realisation is then added to the pool of data and a new location is then visited and the process repeated until a realisation of the full surface is generated. *Sequential Gaussian simulation (SGS)* requires assuming a normal distribution for the ccdf and is most commonly implemented. *Indicator Simulation (IS)* is based on similar approaches to Indicator Kriging (IK) that discretises the ccdf, rather than making distributional assumptions.

3. *Strategies based on optimisation.* These strategies begin with an initial image which is then modified until target statistics (such as the variance or variogram) are met. An optimisation approach, such as simulated annealing, is used for this (Atkinson 1999b, Dungan 1999).

Sequential approaches are now, perhaps, the most widely used strategies for simulation (Deutsch & Journel 1998, Heuvelink 1998, Atkinson 1999b, Dungan 1999, Berterretche et al. 2005), although methods based on optimisation are also gaining interest (Atkinson 1999b).

It is usual to simulate multiple realisations of a given surface. These can be used in the contexts described above (Journel 1996a, Journel 1996b, Heuvelink 1998, Deutsch & Journel 1998, Atkinson 1999b, Dungan 1999, Berterretche et al. 2005). One possible use of simulated surfaces advocated by Dungan (1999) (described above) was as an input to physically-based or empirical-statistical models. However, it should be noted that the error that is added back to the kriged prediction to create the simulated surface is *orthogonal* (i.e., unrelated) to the predicted variable (Curran & Atkinson 1998, Deutsch & Journel 1998, Atkinson 1999b). Hence, Atkinson (Curran & Atkinson 1998, Atkinson 1999b) recommends that conditionally simulated surfaces should *not* be used in regression, since the estimated parameters will differ from those obtained for the original data. Instead they recommend applying an adjustment factor to the regression coefficients to adjust for the variability lost through smoothing (Atkinson & Kelly 1997, Atkinson & Tate 2000). However, whilst this approach was shown to increase the accuracy of parameter estimation it was not conclusive. Hence, conditionally simulated surfaces will still be examined in this thesis.

Finally, it may be noted that simulation algorithms have traditionally worked with the residual component of equation 2.69 (Pebesma 2003) and hence can be seen as the simulation equivalents of SK (see also: Deutsch & Journel 1998, Dungan 1999). However, it is also desirable to account for uncertainty in β, as shown in equation 2.14 and discussed in the surrounding text (Pebesma 2003).

2.3 Positional uncertainty

All previous discussion has made the assumption that, when spatial analysis is performed, the location of those measurements is known without error. However, knowledge of location is central to the implementation of geostatistical and other forms of spatial analysis. It is also central to the implementation of regression where remotely sensed data and field measurements are combined in empirical models, since the field data must be combined with their spatially coincident remotely sensed data. The latter class of models includes the empirical line model for atmospheric correction, which is examined in this thesis. There are numerous other examples in remote sensing and GIS. For example, the objective might be to predict vegetation properties, such as green leaf area index (GLAI) (Curran & Williamson 1985) or biomass (Gao 2006). The development of such models is contingent on being able to position accurately the field measurements within the image. However, despite its obvious importance, the likely effect of positional uncertainty has received only limited mention in the literature (e.g., Justice & Townshend 1981, Prince & Astle 1986, Larsson 1993, Oudemans et al. 2002, Gao 2006). Furthermore, examples of attempts to formalise and quantify this are few (Chiles 1976, Atkinson 1996, Chiles & Delfiner 1999, Salvador 1999, Gabrosek 1999, Gabrosek & Cressie 2001, Gabrosek & Cressie 2002, Cressie & Kornak 2003)

In both the geostatistical and regression case error in locational information typically leads to error in attribute information which ultimately affects estimation and prediction (Gabrosek & Cressie 2001, Gao 2006). It is essential to distinguish between attribute uncertainty which is due to positional uncertainty and that which is not (such as measurement error). This is because the attribute error that arises due to positional uncertainty will be cross-correlated with the underlying variables rather than being orthogonal to it (Atkinson et al. 1996, Atkinson 1996, Cressie & Kornak 2003). This holds even when the positional errors themselves have no spatial structure (Cressie & Kornak 2003).

Consider a variable of interest, $y(\mathbf{u})$. If there is error in $\mathbf{u}$ then the *observed* value of y, y^* will be

$$y^* = y + \delta \tag{2.99}$$

where δ is the attribute error associated with the positional uncertainty. Alternatively, this might be expressed as:

$$y(\mathbf{u}^*) = y(\mathbf{u} + \zeta) \tag{2.100}$$

where ζ is the positional error. Following on from equation 2.99 it can be shown that:

$$\sigma^2_{y^*} = \sigma^2_{(y+\delta)} = \sigma^2_y + \sigma^2_\delta + 2\sigma_{y\delta} \tag{2.101}$$

where σ^2_y indicates the variance of y and $\sigma_{y\delta}$ indicates the covariance between y and δ. However, since both y and y^* are realisations of the same underlying population it is

expected that $\sigma_y^2 = \sigma_{y^*}^2$. Hence

$$\sigma_\delta^2 = -2\sigma_{y\delta} \tag{2.102}$$

Hence there is a negative covariance between y and δ. This result is demonstrated independently by Atkinson (1996) and Salvador (1999). Importantly, the fact that $\sigma_y^2 = \sigma_{y^*}^2$ means that positional uncertainty should make no difference to the sample variance and, for bounded variograms, the variogram sill. The relationship shown in equation 2.101 can also be expressed in terms of variograms:

$$\gamma_{y^*}(\mathbf{h}) = \gamma_y(\mathbf{h}) + \gamma_\delta(\mathbf{h}) + 2\gamma_{y\delta}(\mathbf{h}) \tag{2.103}$$

where γ_y is the cross variogram between y and δ and h is the lag. Hence, for $\mathbf{h} < \phi$ (the range of spatial variation), $\gamma_{y^*}(\mathbf{h}) \neq \gamma_y(\mathbf{h})$; indeed it is expected that $\gamma_{y^*}(\mathbf{h}) > \gamma_y(\mathbf{h})$. In other words, in the presence of positional uncertainty, it is to be expected that the variogram will be over-estimated for $\mathbf{h} < \phi$. Hence, the apparent strength of the spatial structure will be reduced. This result is demonstrated for both simulated (Atkinson 1996, Gabrosek & Cressie 2002) and real data (Chiles 1976). Furthermore, it has been demonstrated that this bias in the variogram leads to reduced accuracy when kriging is implemented. Specifically, it tends to lead to negative bias and an increase in the mean square prediction error (MSPE) for kriged predictions (Gabrosek & Cressie 2002). Indeed, in simulations, where the positional uncertainty is large, this bias was 15 to 50% and the MSPE 10 to 30% larger than the case where the positional uncertainty was explicitly modelled. The lowest accuracy was observed when σ_δ^2 was large relative to ϕ (Gabrosek & Cressie 2002).

In the regression case, intuition suggests that positional uncertainty will lead to a decrease in the correlation between the remotely sensed and field variables and that this would be associated with a flattening of the regression line. This would be expected since, in effect, the field data are being paired with the "wrong" pixels. This expectation is demonstrated by (Salvador 1999), who models the relationship between tree coverage (TC) and simple ratio (SR = Reflectance (NIR) / Reflectance (Red)) using simple linear regression. In the presence of positional uncertainty they find that $\hat{\beta}_0 = 4.59$, $\hat{\beta}_1 = 2.18$ and $R = 0.41$, whereas after correction for positional uncertainty, $\hat{\beta}_0 = 4.09$, $\hat{\beta}_1 = 3.11$ and $R = 0.59$. In this case β_0 and β_1 are the regression coefficients and R is the correlation coefficient.

Gabrosek & Cressie (2002) postulate two models for positional uncertainty, which can be illustrated with respect to figure 2.5. The first of these is the *design model* where the sampling strategy (represented by the open squares) is fixed in advance but is realised imperfectly. One possible realisation is represented by the filled squares. This imperfection might be due to the use of a positioning device (such as a GPS) that is not perfectly accurate or because of the difficulty of precisely positioning the measurement equipment. If it is assumed that the positional uncertainty is unbiased, then the relationship between

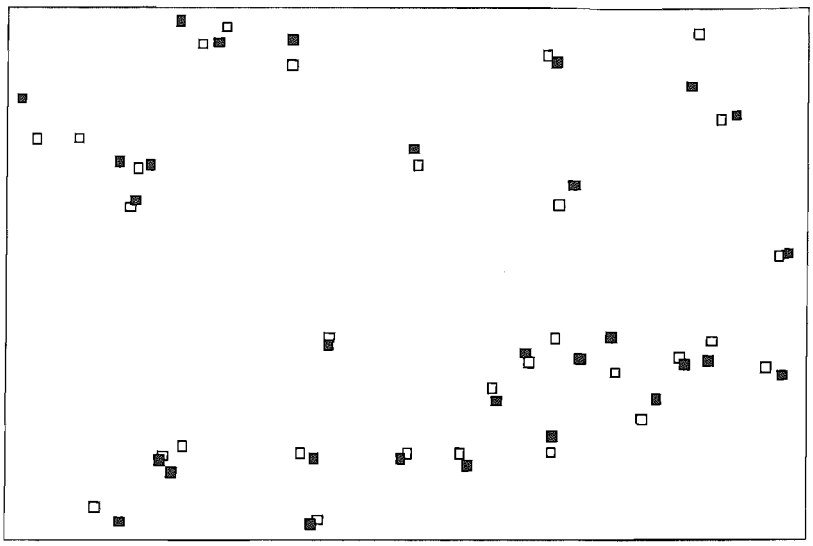


FIGURE 2.5: Plot showing intended locations (open squares) and actual realised locations (filled squares) of field-measurements.

the realised sites, R, and the intended sites, S, might be given as :

$$R|S \sim N(S, \Sigma)$$

whereby the error is assumed to be unbiased and the realised sites are distributed around the intended site according to a probability distribution with mean S and covariance Σ. Hence the attribute is measured at R (which is unknown), but attributed to S (which is known). The emergent effect is that the value attributed to location S is given with error. The design model is referred to as the *coordinate-positioning model* (CP) by Cressie & Kornak (2003). The naming convention of Cressie & Kornak (2003) is adopted in this thesis to avoid confusion with design-based sampling.

The second case is termed the *resource model*. In this situation, a particular resource (or set of resources) at location A is the object of investigation. An example might be the sampling of trees in a forest. The attributes of that resource are measured but the location is measured imperfectly and assigned to location B. Again, assuming that the error in location is unbiased, the model is given as:

$$B|A \sim N(A, \Gamma)$$

Hence the reported positions are distributed around the actual positions with mean A and covariance Γ. In contrast to the design model, B is what is reported, whereas A is unknown. The resource model is referred to as the *feature-positioning model* (FP) by Cressie & Kornak (2003).

An alternative representation of the CP and FP models is given as follows:

$$x = u + d \tag{2.104}$$

where u is the true value of x. If x is an intended location and u is the realised location then we have the CP model. If u is the actual location but x is the reported location, we have the FP model.

Cressie & Kornak (2003) note that the CP and FP models seem very similar. However, they emphasise one important difference. As stated above, under the CP model the location error is centred around the *known* location. Under the FP model the location error distribution is centred around the *unknown* location. Hence, if the objective is to account for positional uncertainty, it is important to choose the correct model. However, where positional uncertainty is not accounted for, it is argued here that that the two forms of positional uncertainty are interchangable.

Most studies of positional uncertainty adopt the CP model. However, for demonstrating the effects of positional uncertainty, these used *simulated* data (Atkinson 1996, Gabrosek & Cressie 2002), where it is straightforward to simulate such errors. Cressie and co-workers (Gabrosek & Cressie 2001, Gabrosek & Cressie 2002, Cressie & Kornak 2003) incorporate the distribution of the (positional) errors into the estimation of the trend and covariance function, hence correcting for positional uncertainty both in estimation and prediction. They show that this increases the accuracy of estimation and prediction, relative to the case where positional uncertainty is ignored, although greater accuracy can still be achieved by reducing or eliminating positional uncertainty. It should be recognised that these studies were undertaken using simulated data or under conditions where the distribution of ζ (the positional error in equation 2.100) was well understood. Hence it is unclear how their methods would perform in the absence of knowledge of the distribution of the positional error.

2.4 Spatial regression

The previous two sections have discussed regression and spatial analysis under different headings. However, in many cases models for spatial analysis can be considered as extensions of the linear regression models. In particular, section 2.2.2.4 illustrated how geostatistics could be used to solve the problem of correlated residuals that was presented in section 2.1.4. Indeed, statisticians generally consider geostatistics to be an extension of the linear model (Ver Hoef et al. 2001, Diggle et al. 2003, Hengl et al. 2003).

Despite the clear connection between geostatistics and regression, there are often important differences in terms of outlook and application (e.g., Dungan 1998, Berterretche et al. 2005). It has already been mentioned that much geostatistical theory and practice was, at least in the early stages, developed separately to mainstream statistics. This point will not be further explored here. However, the differences in application are important. In particular, in geostatistics the focus is generally on understanding the nature of the spatial variation and in prediction, where prediction takes the form of spatial interpolation. In

regression the focus is generally on understanding the relationship of y with $\mathbf{X}$ and in prediction, which might be considered as interpolation within the (spatial) sample space of $\mathbf{X}$. For example, in the case of KT (as presented by Goovaerts 1997, Goovaerts 1999, Webster & Oliver 2001), the form of the trend is modelled within a local neighbourhood using a limited number of data (perhaps as few as 5). As will be shown in chapter 5, this may be unsatisfactory if the objective is to accurately model β. However, they may be appropriate in the context of accurate *spatial* prediction. In the regression context it has been demonstrated that it is necessary to account for spatial correlation in regression in order to identify appropriate and scientifically meaningful covariates and to provide an accurate model of uncertainty for estimation and prediction (Cook & Pocock 1983, Harris & Johnson 1996, Lark 2000, Ver Hoef et al. 2001, Lark & Cullis 2004, Lark & Webster 2006).

This final issue arises because estimation of β is dependent, in the most general case, on the covariance matrix. Hence:

$$\hat{\beta} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

If the residuals are homoskedastic, but have a correlation structure, then $\boldsymbol{\Sigma}$ in equation 2.105 can be replaced with the correlation matrix, $\mathbf{R}$, or a modified correlation matrix, as in equation 2.97. If the residuals are heteroskedastic, but there is no correlation structure, then $\boldsymbol{\Sigma}$ can be replaced with a diagonal weights matrix $\mathbf{W}$, as in equation 2.32 for weighted least squares (WLS). In many cases, the estimates of β will not be substantially different from those obtained from OLS (Pebesma & Wesseling 1998, Lark 2000, Hengl et al. 2004). However, the possibility of substantial and important differences cannot be ruled out (Harris & Johnson 1996, Ver Hoef et al. 2001, Lark & Cullis 2004).

In addition to the point estimators of β, the correlation and weighting structure also influences the estimation of σ^2 as illustrated in equation 2.84 and equation 2.33. Furthermore, $C(\hat{\beta})$, which describes uncertainty in the estimate of β, is also dependent on the covariance structure, as illustrated below:

$$C(\hat{\beta}) = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \tag{2.105}$$

Hence, even if the GLS estimator of β is similar to the OLS estimator, $C(\hat{\beta})$ will differ. That has implications for quantifying uncertainty in estimation and prediction, for specifying confidence and prediction intervals and for hypothesis testing. In particular, the OLS estimators tend to yield confidence intervals that are too narrow (Harris & Johnson 1996, Ver Hoef et al. 2001) and can affect the outcome of hypothesis testing (Lark & Cullis 2004). If correlation is not accounted for then it can be conflated with a trend, leading to the inclusion of spurious covariates in a regression model (Ver Hoef et al. 2001, Hengl et al. 2004). Hence, even if sparseness of data makes spatial interpolation impractical for prediction, it is still necessary to account for spatial autocorrelation at the parameter estimation stage. These issues are given further attention in chapter 7.

There are two further theoretical issues which merit discussion, although they will not be explored in detail in this thesis. First, it should be noted that, whilst proper accounting for the covariance structure should lead to an improved model with more accurate parameter estimates, there may still be problems when it comes to hypothesis testing. This is because the specification of critical values for the t-distribution is dependent on the degrees of freedom. In the presence of autocorrelation, it is not clear how the degrees of freedom should be specified (Harris & Johnson 1996) and this does not seem to be taken into account when calculating critical values (Ver Hoef et al. 2001, Lark & Cullis 2004). Examination of the appropriate statistical tables (Neter et al. 1996) shows that the change in critical value as the degrees of freedom decrease is largest for small samples. Hence, for large samples, this issue is likely to be less significant. Second, Harris & Johnson (1996) note that, where ϕ needs to be estimated, this can lead to problems with the assumptions underlying GLS. In particular, the distributional assumptions for $\hat{\beta}$ and $\hat{\sigma}^2$ may not be met. However, they do not provide guidance for dealing with this. The results presented in chapter 7 explore the relation between correlation structure and estimation of these parameters, although a general theoretical solution is not offered. It might be argued that the issues presented in this paragraph may be solved by moving to the Bayesian approach. This allows for full incorporation of uncertainty at all stages of the estimation and prediction process. Distributions can be constructed using simulation and confidence and prediction intervals similarly computed. However, such things are more difficult to implement in practice and this is deferred as a subject for future research.

This section has emphasised the importance of properly accounting for spatial correlation in the residuals when implementing a regression model. This is important even if prediction by spatial interpolation is not the ultimate goal.

2.5 Summary

This chapter has provided a detailed overview of the relevant statistical methods that will be applied in this thesis. In particular regression and geostatistical techniques were discussed as well as various other issues concerning the use of spatial data. In particular the following issues were raised:

- When utilising regression models with spatial data it is unlikely that the regression requirement of *independent and identically distributed* (iid) residuals will be met. If the residuals are not identically distributed this needs to accounted for using a weighting structure. If they are correlated (and hence not independent), then the correlation structure needs to be accounted for.
- It is generally assumed that there is no error in the predictor variable, x. However, this is not always the case. If information on the relative uncertainty in x and y

is available, then this can be incorporated using the approach of Kendall & Stuart (1967). This is applied in chapter 4.

- The spatial support of the data and model output are important considerations. In particular it was emphasised that care should be taken to match the support when two or more measurements are combined in empirical or physical models.

- It was shown that geostatistical approaches could be used to model the spatial correlation of regression residuals. From a statistical perspective, this works as an extension of the linear model. However, there are practical differences and epistemological differences between applying regression and geostatistical models. Hence, the approach discussed provides a mechanism for accounting for spatial correlation in residuals in situations where geostatistical interpolation might be inappropriate. An extension of this to cover heteroskedastic residuals is discussed in chapter 7.

- It is generally assumed implicitly that the location of spatial data is known without error. However, this is rarely the case. An overview of this problem was discussed in section 2.3 and is considered in chapter 6.

Finally, it should be recognised that, whilst this thesis focuses on the empirical line method (ELM) for atmospheric correction, the issues raised in this chapter are of much more general importance in remote sensing and GIS. Hence the research presented in this thesis has wider implications than the ELM. This is discussed further in chapter 8.

Chapter 3

Remote sensing: theory and models

The objective of this chapter is to provide an overview of remote sensing methods and models that are relevant to this thesis. Discussion of remote sensing is limited to the optical domain and particularly to reflective remote sensing rather than the thermal infrared region. Particular attention is given to atmospheric effects and atmospheric correction and specifically to the empirical line method for atmospheric correction.

3.1 Physical basis of remote sensing

The objective of this section is to provide a concise outline of the most relevant physical processes and terminology in optical remote sensing. This topic is widely covered in introductory text books (Curran 1985, Campbell 1996, Lillesand et al. 2004) as well as advanced text books (Schott 1997, Richards & Jia 1999, Liang 2004) and edited books (Swain & Davis 1978, Asrar 1989). Hence a complete exposition is not required and detailed derivations are not provided.

Remote sensing is based on making measurements of electromagnetic (EM) radiation upwelling from the surface. It is well known (Silva 1978, Liang 2004) that the radiant exitance, M, from a surface is given by the Stefan-Boltzmann radiation law

$$M = \varepsilon \sigma T^4 \tag{3.1}$$

which has units of Wm^{-2} and where ε is the emissivity of the surface, σ is the Stefan-Boltzmann constant and T is the temperature of the surface in Kelvin. It can then be shown that the wavelength, λ_{max}, at which M_λ is at a maximum is

$$\lambda_{max} = \frac{2898}{T} \tag{3.2}$$

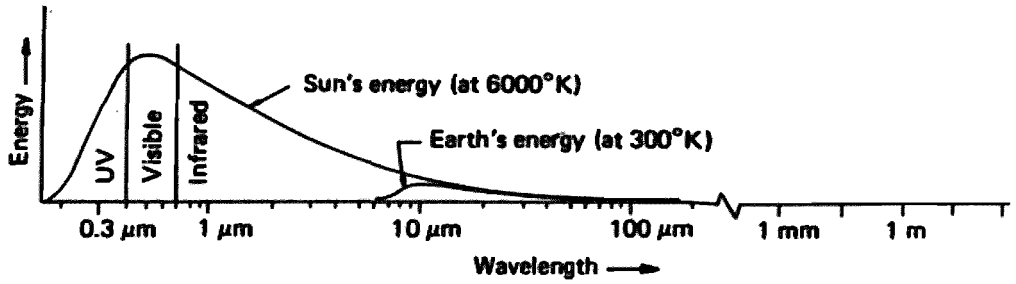


FIGURE 3.1: Diagram showing radiant energy as a function of surface temperature. Source: Lillesand & Kiefer (1994) (p. 5).

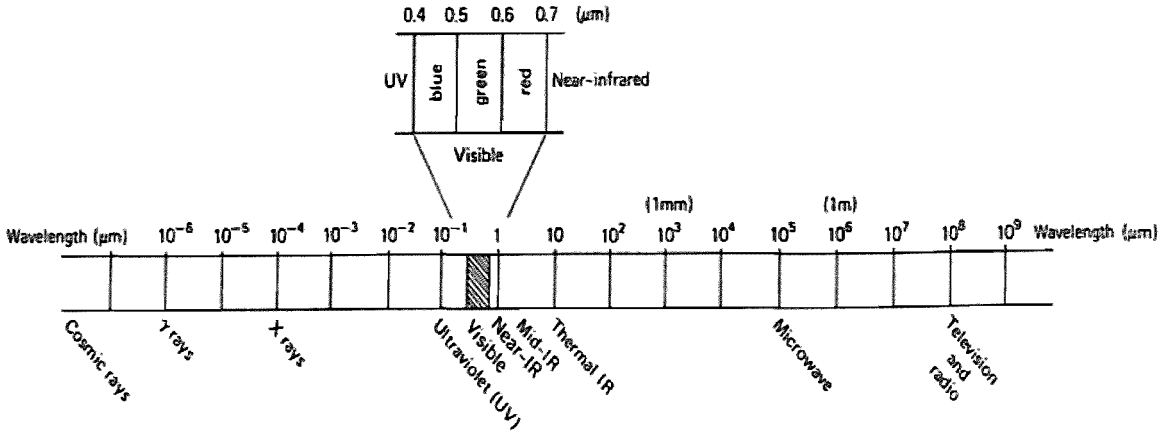


FIGURE 3.2: Diagram showing the different wavelength regions of the electromagnetic spectrum. Source: Lillesand & Kiefer (1994) (p. 11).

which has units of μm. Hence, as the temperature increases, M_λ moves to shorter wavelengths, as illustrated in figure 3.1. Hence, the Sun with a temperature of approximately 6000 K has a peak M_λ at visible wavelengths. However, the Earth, with a typical surface temperature of approximately 300 K has a peak M_λ in the thermal infrared region of the EM spectrum.

The electromagnetic *spectrum* is commonly split into different regions, as shown in figure 3.2. This partitioning is done by wavelength although there is an equivalence between wavelength (λ, which has units of distance, m) and frequency (ν, measured in Hertz, Hz):

$$\lambda = \frac{c}{\nu} \tag{3.3}$$

where c is the speed of light (which has units of speed, m s^{-1}.

The optical region can be further broken down into the reflective region (encompassing the visible, near and middle infrared regions) and the thermal infrared. In the reflective region the *emitted* radiant exitance is minimal and radiation upwelling from the surface will be that reflected from the surface. However, in the thermal infrared region the signal will be comprised of reflected and emitted radiation.

In the absence of atmospheric effects (discussed in section 3.1.1) radiation incident on the surface may be *reflected* (ρ), *absorbed* or *transmitted*. Radiation upwelling from the surface will be comprised of that reflected from the surface and, for longer wavelengths emitted from the surface. Natural surfaces are likely to be complex and the radiation pathway will be complicated by multiple scattering. For example, radiation may be reflected from a leaf and then re-reflected from another leaf or it may be transmitted through a leaf and reflected off a trunk. However, with each scattering event, the radiant intensity will be decreased.

The radiance upwelling from the surface is measured by the sensor. This might be a hand-held (or mounted) radiometer (multispectral) or spectrometer (hyperspectral). Note that the hand-held radiometer may also be mounted on a tripod or mast, but in both cases will, most likely, be directly operated by the user. An imaging radiometer or spectrometer is normally mounted on an airborne or spaceborne platform. It utilises a whiskbroom scanner or CCD array, combined with the forward motion of the platform to build up an image (Campbell 1996).

The area imaged by the sensor is determined by the height of the sensor above the surface and the angular Field of View (FOV) of the sensor. The FOV can also be expressed as an area, which is height dependent. The intensity of the signal within the FOV is also dependent on the lens optics and sensor characteristics and can be represented by the point spread function (PSF) (Schott 1997). For imaging spectrometers, the situation is complicated because the measurements are typically resampled in order to construct an image with square pixels. Hence, the pixel is only an approximation of the areal FOV and will include some signal gathered from outside the geometric confines of the pixel (Schott 1997, Liang 2004) and may even be a convolution of several FOVs. For a hand-held radiometer the angular FOV is typically narrow and is often approximated as representing a point rather than an area.

3.1.1 Atmospheric effects

Under field conditions the atmosphere has highly significant effects, which are reviewed here. Indeed, the atmosphere can be conceptualised in much the same way as any other surface. For modelling purposes, transmission and emittance are generally assumed to be negligible (atmospheric constituents are assumed to be opaque). Hence the key processes are reflectance (scattering) and absorption (Kaufman 1989, Liang 2004).

Figure 3.4 shows the components of upward radiance, L'. The prime indicates upwards radiance whereas L represents downward radiance. In mathematical terms, L' is given as (Kaufman 1989, p. 354):

$$L' = L'_p + L'_s + L'_{d1} + L'_{d2} \tag{3.4}$$

where:

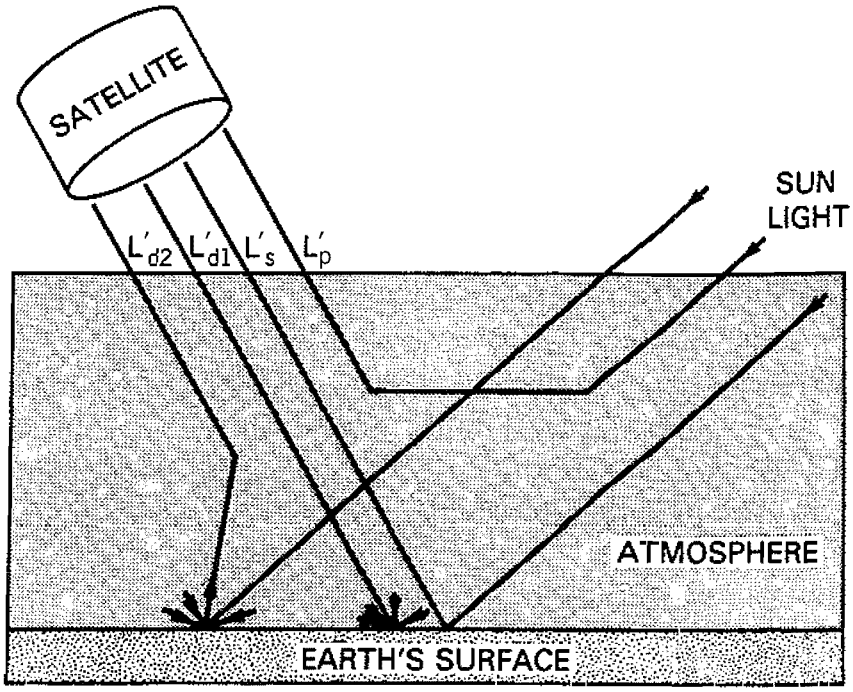


FIGURE 3.3: Schematic presentation of the components of upward radiance. Acronyms explained in text (below). Source: Kaufman (1989, p. 354).

L'_p is the path radiance, which is scattered from the direct sunbeam by the atmosphere into the sensors field of view without ever reaching the surface. This component tends to uniformly brighten the image, resulting in a loss of contrast.

L'_s has been transmitted directly downwards through the atmosphere and reflected by the surface directly into the sensor.

L'_{d1} is the diffuse irradiance that is reflected directly back to the sensor. This diffuse illumination arises due to scattering from the solar beam by the atmosphere. It may also arise if the incident beam reflects off other surfaces *prior* to being reflected into the sensor path.

L'_{d2} is the component that is reflected from neighbouring surfaces and then scattered by the atmosphere into the sensor path. This effect tends to be particularly significant for fine spatial resolution sensors that are imaging high contrast surfaces.

The latter three effects are influenced by properties of the target as well as the atmospheric properties. Hence they will be affected by the spatial uniformity of the surface and by whether the surface exhibits Lambertian properties.

As mentioned above, the two key atmospheric processes affecting the signal are *absorption* and *scattering* (Silva 1978, Kaufman 1989, Liang 2004). Absorption attenuates the signal by reducing the amount of energy available at any given wavelength. Scattering changes the direction of radiation and, hence, tends to reduce the contrast and cause blurring of the image.

Ultraviolet and visible	0.30-0.75 μm
Near infrared	0.77-0.91 μm
Mid infrared	1.55-1.75 μm
	2.05-2.40 μm
Thermal infrared	8.00-9.20 μm
	10.2-12.4 μm
Microwave	7.5-11.5 mm
	20.0+ mm

TABLE 3.1: Major atmospheric absorption windows. Source: Campbell (1996).

There are three key components of the atmosphere that cause absorption and scattering: *gases*, *aerosols* and *clouds*. The main components of atmospheric gases are nitrogen (N_2: 78 %) and oxygen (O_2: 21 %) although there are small amounts of water vapour (H_20), carbon dioxide (CO_2) and ozone (O_3). Carbon dioxide tends to be stable and well mixed in the atmosphere, although concentrations may be higher above large sources, such as cities. Ozone is concentrated in the stratosphere (20 - 50 km above the surface) and its distribution is well understood. Water vapour is found mainly in the boundary layer (1 - 2 km), but is highly variable both spatially and temporally (Kaufman 1989). The variability in water vapour poses a major challenge to quantifying its effects on scattering and absorption (Kaufman 1989, Liang 2004).

Atmospheric gases have clearly defined *absorption bands*. It is these absorption bands that lead to *atmospheric windows* (See table 3.1) where remote sensing can be performed. Outside these atmospheric windows the signal is strongly attenuated. As discussed above, the distribution of most atmospheric gases is consistent and well understood. The major exception is water vapour, which needs to be measured or predicted. However, it is possible to make use of the absorption bands to estimate atmospheric constituents. In particular, the *depth* and *width* water absorption bands at 840 nm, 930 nm and 1130 nm can be used to predict atmospheric water vapour (Gao & Goetz 1990, Goetz et al. 2003). For hyperspectral sensors, this approach can be used to predict water vapour on a pixel-by-pixel basis.

Scattering by gases is referred to as molecular scattering or *Rayleigh scattering* (attributed to the 19th century British physicist J.W. Strutt (1871)[1] (Kaufman 1989)). Rayleigh scattering is caused by matter that is much smaller than the wavelength of energy being scattered. Its effect is most pronounced at short wavelengths – particularly ultraviolet and visible wavelengths. Rayleigh scattering is the reason why the sky appears blue most of the day. The sky may also appear red at sunrise and sunset where the atmospheric path length is greater, since the sun is on the horizon. The principles of Rayleigh scattering are well understood. It is dependent on the Rayleigh optical depth (τ_R), which is modelled as a function of wavelength and pressure (Kaufman 1989, Liang 2004).

[1]The Honourable J. W. Strutt became the 3rd Baron (Lord) Rayleigh in following the death of his father in 1873.

Atmospheric aerosols are liquid or solid matter suspended in the air. Liquid particles greater than 1 μm are usually referred to as cloud drops (Kaufman 1989). Aerosols may be formed through the weathering of solid surfaces and the removal of small particles by the wind, as a result of fire or due to the bursting of bubbles at the ocean surface. Aerosols may also form in the atmosphere due to the aggregation of gaseous particles and the condensation of gases onto existing particles. The aerosol composition of the atmosphere varies in time and space. Furthermore, the size distribution of atmospheric constituents also varies. However, "standard" atmospheres are available for different regions of the world which may serve as a reasonable approximation in any given situation.

Scattering by aerosols may be by *Mie scattering* (attributed to the German physicist Gustav Mie (1908) (Kaufman 1989)) or *non-selective scattering*. Mie scattering is relevant where the aerosols are similar in size to the wavelength of the energy being scattered, whereas non-selective scattering is relevant where the aerosols are much larger. The optical effects of aerosols depend on the physical characteristics of the particles, specifically the size and shape distribution, the spatial variation in aerosol concentration and the refractive index of the particles (Kaufman 1989). Measuring these atmospheric properties is difficult, so it is common practice to use approximations based on geographic location, length of season and atmospheric humidity (Kaufman 1989) and databases are available to support this (Liang 2004). Mie code can then be used to yield the optical depth (τ), single scattering albedo (ω) and the scattering phase function($P(\cdot)$) (Kaufman 1989, Liang 2004).

Finally, clouds are an important atmospheric constituent. Large clouds that completely fill a pixel will be apparent in a remotely sensed image because, in the optical domain, they obscure the ground. They may also cast shadows on the ground. However, for smaller clouds the effects may be less obvious especially if they do not completely fill pixels. There are algorithms available to mask out cloud contaminated pixels (Campbell 1996). If known the optical characteristics of clouds can be described by the pixel cloud fraction, the cloud size distribution and the reflection and transmission properties (Kaufman 1989). Owing to the problems posed by clouds it is common practice to limit airborne and field campaigns to cloud-free conditions (Milton 1987). However, this is not always possible, particularly in temperate environments (such as southern England). Furthermore, Milton & Goetz (1997) find that the diffuse to global ratio can change over short time-scales due to changes in water vapour and high-altitude cirrus clouds that may not be apparent to the human observer.

The optical characteristics of a small volume of atmosphere can be described by three key parameters (Kaufman 1989, p. 346):

K_e, the extinction coefficient (m^{-1}), which represents the attenuation of radiation propagating through the atmosphere. The integral of K_e over altitude gives the more commonly quoted optical thickness (or optical depth), τ. The aerosol optical thickness depends on the aerosol characteristics and the total aerosol loading.

ω, the single scattering albedo, which is the ratio between the scattering coefficient and the total extinction coefficient (absorption and scattering). It is important to note that this accounts for *aerosol* absorption (Kaufman 1989, Liang 2004).

$P(\theta)$, the scattering phase function which describes the angular distribution of scattered radiation. It is dependent on particle size and optical thickness.

The optical depth, single scattering albedo and scattering phase function describe the characteristics of a small volume of atmosphere. However, this then needs to be integrated over the entire depth of the atmosphere. This is performed through the radiative-transfer (RT) equations, attributed to Chandrasekhar (1950) (Kaufman 1989, Liang 2004).

The RT equation traces the change in radiance ($L'(\theta')$) through a thin vertical layer of atmosphere, where θ' is the zenith angle of L'. The vertical component (z) is generally replaced with the optical thickness (τ) (Kaufman 1989, Liang 2004). The RT equation is given as (Kaufman 1989):

$$dL' = -L'\frac{d\tau}{\cos\theta'} + (J + J_o)\frac{d\tau}{\cos\theta'} \tag{3.5}$$

The first term on the RHS represents the *attenuation* of L' due to gaseous absorption in the atmospheric layer. This increases exponentially with optical thickness, according to Beer's law (Liang 2004). The second term on the right represents the increase in radiance due to scattering of direct sunlight and diffuse illumination. Here, $J_o = f(\omega, \tau, P, E_s)$ is a source term representing the scattering of direct sunlight and is a function of the ω, τ, P and E_s (the extraterrestrial solar irradiance). The other source term, $J = f(\omega, P, L')$ is integrated over all zenith and azimuth angles. It is due to diffuse skylight and multiple scattering between the atmosphere and the surface. Hence J is also dependent on the surface reflectance properties. The RT equations can be solved numerically or using faster (but less accurate) approximations (Vermote et al. 1997, Liang 2004).

3.1.2 Nomenclature for reflectance

The discussion in section 3.1 introduced the concept of reflectance in general terms. However, reflectance is not a straight forward concept since it is dependent on the viewing and illumination conditions. The objective of this section is to introduce the concept of directional reflectance and clarify the terminology that will be used in the remainder of the thesis. Table 3.2 provides a list of radiometric terms.

The concept of directional reflectance is illustrated in figure 3.4. In the absence of atmospheric scattering, irradiance will come from a collimated beam (direct illumination). The broad downwards arrows represent illumination attributed to (i) radiation that has been reflected from or transmitted through other elements of the surface (consider the tree example given above) or (ii) diffuse illumination from the atmosphere (Martonchik et al. 2000).

Symbol	Description	Defining Expression	Units	Units (Abbreviation)
Q	Radiant energy	Q	Joule	J
Φ	Radiant flux (power)	dQ/dt	Watt	W
M	Radiant exitance	$d\Phi/dA$ (out)	Watts per square metre	W m^{-2}
E	Irradiance	$d\Phi/dA$ (in)	Watts per square metre	W m^{-2}
L	Radiance	$d^2\Phi/\cos\theta d\omega$ (out)	Watts per square metre per steradian	W m^{-2} sr^{-1}
M_λ	Spectral radiant exitance	$dM/d\lambda$	Watts per square metre per micrometre	W m^{-2} μm^{-1}
E_λ	Spectral irradiance	$dE/d\lambda$	Watts per square metre per micrometre	W m^{-2} μm^{-1}
L_λ	Spectral radiance	$dL/d\lambda$	Watts per square metre per steradian per micrometre	W m^{-2} sr^{-1} μm^{-1}

TABLE 3.2: Radiometric terms for remote sensing. Adapted from Silva (1978) and Nicodemus et al. (1977).

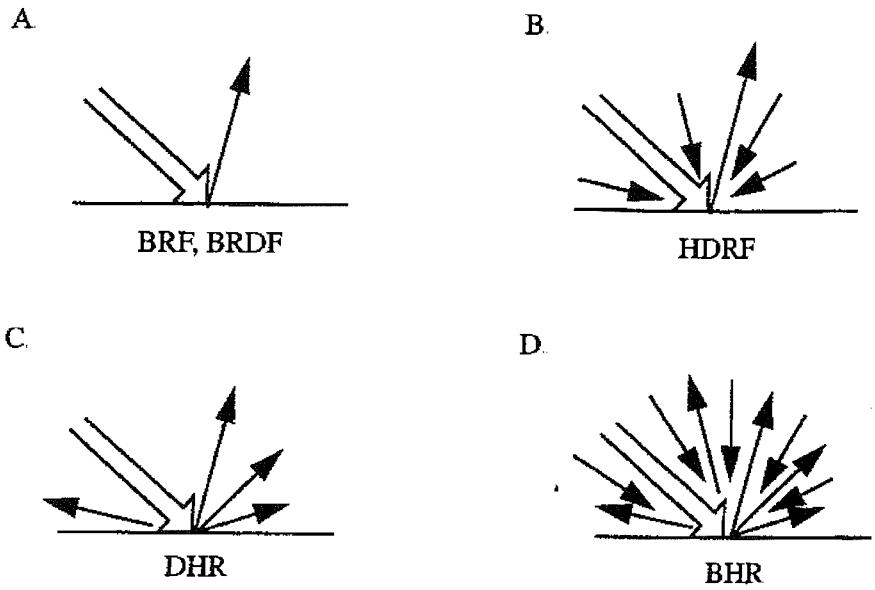


FIGURE 3.4: Diagrammatic summary of reflectance nomenclature. The broad arrow represents irradiance from a collimated beam. All other arrows represent reflected radiation. A. BRF (bidirectional reflectance factor), BRDF (bidirectional reflectance distribution function). B. HDRF (hemispherical-directional reflectance factor). C. DHR (directional-hemispherical reflectance factor). D. BHR (bi-hemispherical reflectance factor). Source: Martonchik et al. (2000, p. 14).

Standards for directional reflectance nomenclature are generally traced back to (Nicodemus et al. 1977), although several more recent reviews are also useful (Milton 1987, Martonchik et al. 2000, Raffy & Blamont 2003, Schaepman-Strub et al. 2006). It is standard practice to quote the angular characteristics of the illumination first, followed by the reflected radiance. Hence "hemispherical directional" implies hemispherical illumination with reflectance measured in a specific direction (Martonchik et al. 2000).

3.1.2.1 Bidirectional reflectance distribution function (BRDF)

Consider a situation where there is no atmospheric scattering, as shown in figure 3.4. The energy from the Sun and the energy reflected to the sensor can be thought of as being confined to two slender elongated cones, each subtending a small angle at the target surface (the solid angle, measured in units of steradians (sr)) (Schott 1997). If these solid angles are sufficiently small, the reflectance is defined in terms of the *bidirectional reflectance distribution function* (BRDF) as (Nicodemus et al. 1977, Martonchik et al. 2000):

$$BRDF(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{dL_r(\theta_i, \phi_i; \theta_r, \phi_r)}{dE(\theta_i, \phi_i)} \mathrm{sr}^{-1} \tag{3.6}$$

with units of steradians (SR^{-1} where:

dL_r is the reflected radiance per unit solid angle;

dE is the irradiance

θ and ϕ denote zenith and azimuth angles, respectively;

i and r denote incident and reflected rays respectively.

and where dependence on wavelength is suppressed. To fully characterise the BRDF it is necessary to measure dL and dE at all possible viewing and illumination positions. However, in experimental conditions it is not possible to measure dE at the target surface (Milton 1987). Furthermore it is defined over differential solid angles and is, hence, theoretically impossible to measure (although it may be approximated) (Martonchik et al. 2000).

3.1.2.2 Bidirectional reflectance factor (BRF)

The *bidirectional reflectance factor* (BRF) is provided to give a measurable alternative the BRDF (Milton 1987, Milton et al. 1995, Martonchik et al. 2000). In this context, the term "bidirectional" refers to the two angles involved (i.e., those defining the source and sensor positions). It is found by standardisation of reflected radiance by the reflected radiance of a panel which is specified to be perfectly diffuse and completely reflecting. It is viewed under the same viewing and illumination conditions as the target (Silva 1978). This measurement configuration is termed *bi-conical* since both the target and panel are sensed using instruments with a finite narrow, conical field of view (Milton 1987). The BRF is given as (Martonchik et al. 2000):

$$BRF(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{dL_r(\theta_i, \phi_i; \theta_r, \phi_r)}{dL_{r_p}(\theta_i, \phi_i)} \tag{3.7}$$

where:

dL_r is the radiance reflected from the surface;

dL_{r_p} is the radiance reflected from a Lambertian panel;

all other terms are as defined for equation 3.6.

The BRF is directly related to the BRDF as (Milton 1987, Schott 1997, Liang 2004):

$$BRF(\theta_i, \phi_i; \theta_r, \phi_r) = \pi BRDF(\theta_i, \phi_i; \theta_r, \phi_r) \tag{3.8}$$

In practice, perfectly diffuse and completely reflecting reflectance panels are not available, so a correction factor, k, is applied (Milton 1987):

$$BRF(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{dL_r(\theta_i, \phi_i; \theta_r, \phi_r)}{dL_{r_p}(\theta_i, \phi_i; \theta_r, \phi_r)} k(\theta_i, \phi_i; \theta_r, \phi_r) \tag{3.9}$$

The use of the BRF instead of the BRDF to represent the spectral reflectance of targets involves several assumptions, as follows (Milton 1987)

1. the field–of– view of the sensor is small (less than approximately 20^o);
2. the reflectance panel must fill the field–of–view of the sensor;
3. there should be no change in the irradiance magnitude or distribution between the measurement of dL_r and dL_{r_p};
4. direct solar flux dominates the irradiance field (i.e. the Sun shines out of a "black" sky);
5. the sensor responds in a linear fashion to changes in radiant flux;
6. the reflectance properties of the standard panel are known and invariant over the course of the measurements.

It is generally possible to ensure that most of these assumptions are met or the effects are limited. However, in field situations the Sun does not shine out of a black sky and the influence of skylight may be significant (Milton 1981a, Duggin & Philpson 1982, Milton et al. 1995). Hence, the BRF cannot, theoretically, be measured in field conditions, although it may be approximated.

3.1.2.3 Hemispherical-directional reflectance factor (HDRF)

The *hemispherical-directional reflectance factor* (HDRF) provides a definition of reflectance where illumination is from the entire upper hemisphere. It is given as (Martonchik et al. 2000):

$$HDRF(\theta_r, \phi_r) = \frac{dL_r(\theta_r, \phi_r)}{dL_{r_p}} \tag{3.10}$$

Hence this formulation is similar to equation 3.7, except that dependence on (θ_i, ϕ_i) has been eliminated. Furthermore, whereas the BRF and BRDF are intrinsic properties of the surface, the HDRF depends both on those properties and on the illumination conditions. As noted in equation 3.5 the illumination conditions are governed mainly by the atmospheric properties, but are also dependent on the surface properties owing to multiple scattering between the atmosphere and the surface (Kaufman 1989, Martonchik et al. 2000, Liang 2004).

The HDRF is the property that is measured in the field or by a remote sensor. Predicting the BRDF from the HDRF requires knowledge of the irradiance distribution which may be obtained through RT modelling (Liang 2004). This modelling is complicated by imperfect knowledge of the atmospheric state and by the fact that the downwelling radiance is dependent on the (unknown) BRDF, owing to multiple scattering between the surface and the atmosphere (Martonchik et al. 2000).

Finally, it may be noted that some authors have used the term *hemispherical-conical reflectance factor* (HRF), rather than the HRDF (Milton 1987). The term conical may be regarded as more appropriate than directional, since it recognises that the reflected radiance is measured over a finite solid angle.

3.1.2.4 Hemispherical reflectance (albedo)

The final terms of interest relate to the *hemispherical reflectance* (Martonchik et al. 2000), also known as albedo (Liang 2004). Previous terms have related to the directional reflectance, whereas the hemispherical reflectance is integrated over all viewing angles. There are two key hemispherical reflectance terms: *directional-hemispherical reflectance* (DHR) and *bi-hemispherical reflectance* (BHR). The DHR is the BRDF integrated over all possible viewing directions. As with the BRDF (and BRF) this is an inherent property of the surface and is not dependence on atmospheric conditions. Conversely, the BHR is the HDRF integrated over all viewing directions. As with the HDRF, this is dependent on diffuse illumination conditions as well as surface-atmosphere interactions. As with the BRDF, the DHR cannot be measured, although it can be predicted if the diffuse illumination conditions and surface-atmosphere interactions are accounted for.

The DHR is sometimes referred to as the "black-sky" albedo. The term "white-sky" albedo is used to refer to the BHR where the sky irradiance is isotropic (Liang et al. 1999, Martonchik et al. 2000). In particular, these terms are associated with MODIS products (Liang et al. 1999).

3.1.2.5 Spectral dependence

In the above discussion, dependence on wavelength (λ) has been suppressed. However, it should be noted that all the quantities discussed are wavelength dependent. Hence, where there is any ambiguity the wavelength or waveband should be stated.

3.2 Atmospheric correction

The physical basis of remote sensing was outlined in section 3.1 with atmospheric effects being covered in section 3.1.1. *Atmospheric correction* is generally taken to mean the prediction of surface reflectance from remotely sensed imagery (Campbell 1996, Liang 2004) and that is the sense that it is meant in this thesis. Schott (1997) calls this process *atmospheric calibration*. This term is not favoured in this thesis, since calibration implies reference to a standard (*Collins Softback English Dictionary* 1992) (Youden 1997, Anderson & Milton 2006). However, it will be shown that atmospheric correction can only be approximately performed and the level of approximation does not merit the term calibration. Finally, it may be noted that the term atmospheric correction may also be used in

other contexts where the atmospheric effects are accounted for, even though the ultimate objective may not be the prediction of at surface reflectance (Kaufman 1989, Martonchik et al. 2000). For example, *relative* correction may be performed where two or more images are corrected against a reference image, but results are not given in absolute units of reflectance.

The radiance measured at the sensor can be given as (Chandrasekhar 1950, Tanre et al. 1979, Deschamps et al. 1983, Caselles & Lopez Garcia 1989):

$$\rho' = L' \frac{\pi}{E_s \cos\theta_i} = \rho_a + \frac{\rho}{1 - \rho S} T_i(\theta_i) T_r(\theta_r) \tag{3.11}$$

where

the subscripts i and r refer to incident and reflected radiation respectively;

ρ' is the apparent reflectance at the sensor;

L' is radiance measured at the sensor (at–sensor radiance);

E_s is exoatmospheric radiance onto the surface;

ρ_a is the reflectance of the atmosphere;

ρ is the surface reflectance;

S is the spherical albedo of the atmosphere;

$T(\theta)$ is the *total* atmospheric transmittance (i.e., *direct* plus *diffuse*);

A similar formulation is given elsewhere (Liang et al. 1997, Liang et al. 2001, Cooley et al. 2002, Liang 2004). Note that ρS is typically a very small quantity, hence $\rho S \approx 0$. Hence, equation 3.11 is often simplified (Moran et al. 1992, Chavez Jr. 1996, Schott 1997), as given in equation 3.12. Note the slightly different definition of T.

$$L' = L_u + (E'_s + L_d) T_r \pi^{-1} \rho \tag{3.12}$$

where:

L' is radiance measured at the sensor (at–sensor radiance);

ρ is the at-surface reflectance factor – this is the quantity that we are interested in deriving;

L_u is the upwelling irradiance along the target–sensor path, also known as the path–radiance;

$E'_s = E_s \cos\theta_i T_i$;

E_s is exoatmospheric radiance onto the surface, perpendicular to the incident beam;

θ_i is the solar zenith angle relative to the surface;

T_i is the atmospheric transmission along the sun–target path;

L_d is the downwelling irradiance (diffuse illumination) onto the target;

T_r is as T_i but along the target–sensor path.

In equation 3.13 dependence on viewing and illumination direction (and wavelength)was suppressed. However, this would, more fully, be given as:

$$L'(\theta_i, \phi_i) = L_u(\theta_i, \phi_i) + (E'_s \pi^{-1} + L_d) T_r(\theta_i, \phi_i) \rho(\theta_i, \phi_i) \tag{3.13}$$

which is conceptually similar to the HRDF. This angular dependence will be suppressed in subsequent discussion.

In order to predict ρ, given L, it is necessary to measure or predict L_u, L_d, T_i, T_r and E_s. For spaceborne sensors E_s is well known, but for airborne sensors this needs to be measured. Procedures for predicting the remaining terms will be discussed in sections 3.2.1, 3.2.2 and 3.2.3.

Liang (2004) distinguishes between methods for correcting *single viewing-angle* and *multiple viewing-angle* imagery. In this context, single viewing-angle imagery refers particularly to sensors which are centred at nadir (e.g., Landsat TM (Thematic Mapper), Airborne TM), although most of the image swath is clearly off-nadir. For such imagery most schemes assume the surface is Lambertian, for the purposes of the atmospheric correction – even if this assumption is not tenable for examination of the surface itself. In contrast, there are several sensors that observe simultaneously the surface from different directions. These include ASAS (Advanced Solid-State Array Spectroradiometer), MISR (Multi-Angle Imaging Spectroradiometer) and Air-MISR. For such sensors, a scheme that accounts explicitly for the BRDF properties of the surface is appropriate (Liang 2004). Single-viewing angle imagery is of primary interest for this thesis.

Schott (1997) discusses only single-viewing angle approaches. He distinguishes three categories:

- *Physically based methods.* These are based on radiative transfer schemes which model the interaction between radiation and the atmosphere.
- *Within image methods.* These techniques are based on data contained within the image itself.
- *Methods based on ground truth: The empirical line method (ELM).* This is based on empirical relationships between measurements of reflectance (made in the field) and measurements of radiance (recorded by the remote sensing instrument).

The distinction between these three categories is not crisp. In particular much research effort is directed at using the imagery itself to obtain the atmospheric variables required for implementation of physically based methods (Gao & Goetz 1990, Liang et al. 1997, Liang 2004).

Physically based methods and within-image methods are discussed briefly below. Further detailed discussion is available in text books (Schott 1997, Liang 2004) and in numerous research papers. Detailed attention is then given to discussion of the ELM.

3.2.1 Physically based methods

Section 3.1.1 discussed the ways in which electromagnetic effects interact with the atmosphere and noted that the RT equations can be used to model those interactions. The theory for RT modelling is well established and numerous software implementations exist (Liang 2004, p. 64-65). However, there are considerable practical difficulties with the implementation concerned with (i) measuring or predicting the necessary atmospheric variables and parameters and (ii) the computational time required to predict reflectance (Liang et al. 1997, Liang 2004).

As discussed in section 3.1.1, particular difficulties are associated with predicting atmospheric water vapour and aerosol loadings and their associated optical depths and scattering coefficients. Information on aerosols can be obtained from Sun photometer and radiosonde data (Shaw 1983, Schott 1997) and information on water vapour is available from a wide range of sources including surface meteorological data, LiDAR, GPS and radiosonde data (Liang 2004). However, such measurements can be expensive to obtain and lack the spatial coverage required to accurately quantify spatial variability (Schott 1997).

The difficulty in measuring the spatial distribution of atmospheric properties has prompted research on predicting these properties from image data. For hyperspectral data (or appropriately positioned multispectral wavebands), the differential absorption technique can be used to predict the total water vapour content based on the depth and breadth of the water absorption bands (e.g., 0.94 μm) (Gao & Goetz 1990, Cooley et al. 2002, Goetz et al. 2002, Goetz et al. 2003, Liang 2004). This approach can be implemented on a pixel-by-pixel basis. Prediction of aerosol optical depth has focused on utilising the correlation in reflectance between the visible and near-infrared (2.1 μm) bands over dark surfaces (Kaufman et al. 1997, Liang et al. 1997). This approach is effective since aerosol scattering is minimal at 2.1 μm but substantial in visible wavebands (Liang 2004). This approach is now operational for MODIS and MERIS sensors (Liang 2004). Information on the single scattering albedo and scattering phase function are available from aerosol climatology data (e.g., Hess et al. 1998) (Liang et al. 2001).

Traditional computational approaches to solving the RT equations have focused on numerical approaches or fast approximations (Liang 2004). Numerical approaches, whilst accurate, can be prohibitively expensive computationally in an operational situation (Liang

et al. 1997, Liang et al. 2001). On the other hand, approximate approaches may prove inaccurate. This has prompted the development of the look-up table (LUT) approach, whereby the computations are performed once and the LUT is queried "off-line" in the operational situation (Liang et al. 2001).

The adoption of physically based approaches is desirable since they may be regarded as the most elegant (Schott 1997) and general (Liang 2004) approaches. However, they are strongly reliant on the ability to measure or predict the relevant atmospheric properties. Substantial advances have been made in techniques to predict those properties – both from within the image and by fusion with other data-sources. However, such approaches may not be suitable in any given situation hence the scenarios in which they can be used remain limited (Schott 1997).

3.2.2 Within image methods

These approaches aim to perform atmospheric correction based primarily on data within the image or within a series of images of the same area. This is, according to Liang (2004) the oldest and most widely used form of atmospheric correction. Furthermore, as discussed in section 3.2.1 this approach is now used to provide inputs to physically based models.

The basis for within-image methods goes back to the *dark object subtraction* (DOS) approach developed by Chavez in the 1970s and 1980s (Chavez Jr. 1988). Recall the equation given in equation 3.12:

$$L' = L_u + (E'_s + L_d)\pi^{-1}T_r\rho$$

For a "dark object" it is postulated that the $\rho = 0$, hence at the sensor $L = L_u$. Hence, the basic DOS approach simply corrected for path radiance and did not provide an absolute measure of reflectance. This approach was refined by Chavez Jr. (1988) by using the DOS approach to broad classification of atmospheric conditions, ranging from "very clear" to "very hazy". This classification is then used to assign an appropriate relative scattering model, which is then applied across all bands. More recent developments (Chavez Jr. 1996) have combined the DOS model with an approximation of T_i (the transmittance on the sun-sensor path) by setting it equal to the cosine of the solar zenith angle ($\cos\theta_i$) (the COST model). This is based on empirical evidence that $\cos\theta \approx \exp(\tau \sec\theta)$, where τ is the optical depth and $T = \exp(\tau \sec\theta)$ is a commonly adopted model for the transmittance. The COST model does not account for L_d and sets $T_r = 1$, since $\cos\theta_r = 0$ for nadir viewing sensors. Chavez Jr. (1996) notes that inaccuracies caused by these approximations are compensated for by inaccuracies caused by setting $T_i = \cos\theta_i$.

Chavez Jr. (1996) evaluated the COST model against the results for physically based models presented by Moran et al. (1992). He found that the COST model delivers a

slightly lower, but acceptable level of accuracy to the physically based models – but does not require access to any field or atmospheric data. Whilst acknowledging the limitations of his evaluation, Chavez Jr. (1996) presents the COST model as an alternative for use in situations where appropriate field and atmospheric data are not available.

The legacy of the DOS approach is very important in atmospheric correction (Liang 2004). As demonstrated in section 3.2.1, developments of this approach have been used in conjunction with other data sets to form an input into physically based models.

3.2.3 Empirical line method (ELM)

The Empirical Line Method (ELM) is a conceptually simple system for converting from at–sensor radiance to at–surface reflectance. One or more *ground targets* (GTs) are selected and the reflectance of the GT is characterised. That characterisation is based on field–sampling, using a radiometer or spectrometer (e.g., Kruse et al. 1990, Smith & Milton 1999, Karpouzli & Malthus 2003) or on laboratory–based measurements of field samples (e.g., Ben-Dor et al. 1994, Ben-Dor & Levin 2000) or some combination of the two (e.g., Farrand et al. 1994). If the mineralogy of the surface can be accurately characterised, then spectral libraries may be used (Farrand et al. 1994). The ELM is based on a simple linear relationship between radiance and reflectance. This is illustrated in figure 3.5 and the physical basis for this relationship is discussed in section 3.2.3.1. The ELM corrects for atmospheric effects and residual instrument artifacts. The ELM can also be used to correct for viewing geometry effects (Kruse et al. 1990) and solar–zenith angle effects (van der Meer 1994).

The ELM has been implemented with one bright GT (e.g., Freemantle et al. 1992, McArdle et al. 1992). This assumes that the darkest point in the image is a "zero reflectance" target (Smith & Milton 1999) and that the radiance measured over that target is attributable to the path radiance (L_u in equation 3.12) (see figure 3.5). The obvious weakness of this approach is that these assumptions may not be met at all or may not be met in all wavebands. Large errors of up to 15% to 20% reflectance (McArdle et al. 1992, Freemantle et al. 1992) have been found when using this approach. This approach has been further developed by Moran and co-workers (Moran et al. 2001, Moran et al. 2003). They advocate devoting substantial effort to accurately characterising the reflectance of a bright GT. The reflectance of a dark target is performed using a RT model using appropriate default atmospheric conditions. This yields an accuracy (mean absolute difference (MAD)) of 1% reflectance for relatively bright targets (Moran et al. 2001). However, it should be noted that these studies have been undertaken at well instrumented, extensive, homogeneous sites in Arizona, USA. Such areas also tend to have stable, clear atmospheric conditions. Hence, their results may not translate well to smaller, less-homogeneous temperate areas of the world. Furthermore, it should be recognised that such results typically apply on a per-target rather than a per-pixel basis – whereas predictions are required on a per-pixel basis. Hence there is a mismatch between the support of the model and the support for

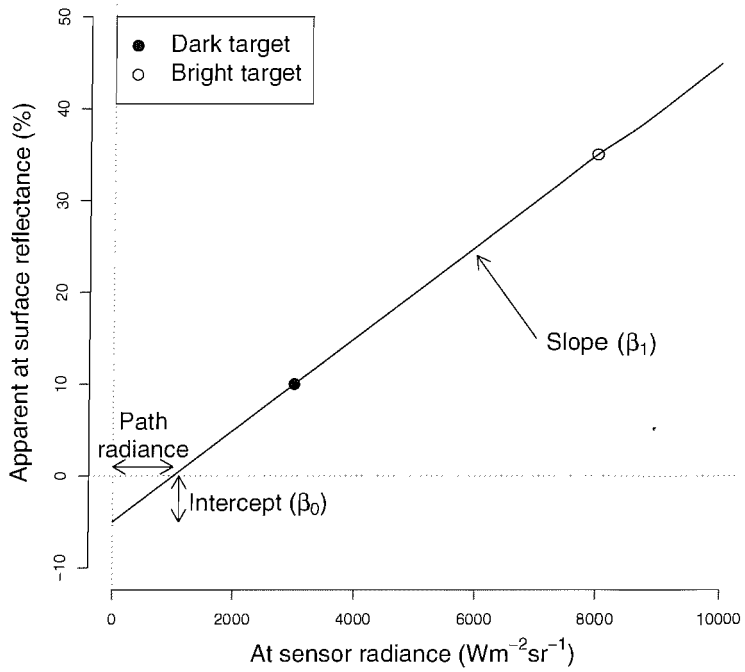


FIGURE 3.5: Schematic representation of the empirical line method.

which the model is required. Hence these statements of accuracy may not apply at the pixel-sized support.

Combined empirical and radiative transfer (RT) schemes have also been implemented for AVIRIS data (Clark et al. 1993, Clark et al. 2000). These provide an empirical correction to the RT scheme and correct for inaccuracies in the RT scheme caused by imperfect models and data. They find this hybrid approach to yield the most accurate correction of AVIRIS data. However, the approach is time-consuming and costly and they estimate that it takes one-person month to accurately characterise a calibration target.

Where the ELM is implemented with two targets, and there is one radiance–reflectance datum for each target, the model parameters (a and b in equation 3.14 may be derived by simply plotting a straight line through the two points on the graph (Farrand et al. 1994, Smith & Milton 1999). This approach obviates the need to undertake regression analysis. However, such an approach is also likely to be sensitive to the characterisation of the bright- and dark-points. Indeed, the error may be as large as 10% (Schott et al. 1988, Caselles & Lopez Garcia 1989).

Where there are more than two targets, or where there is more than one measurement per target, the ELM is implemented using a regression model (Schott 1997, Smith & Milton 1999, Karpouzli & Malthus 2003). This is discussed in detail below.

Finally, it should be noted that the ELM yields an *apparent* at–surface reflectance factor, since, strictly, reflectance is affected by surface properties (including morphology) (Farrand et al. 1994, Ferrier & Wadge 1996), Sun–target–sensor orientation and atmospheric condi-

tions (Kaufman 1989, Farrand et al. 1994, Schott 1997). The strict definition of reflectance is discussed in section 3.1.2. For this thesis, the term reflectance should be taken to mean the apparent at-surface reflectance factor, unless otherwise stated.

3.2.3.1 Physical basis of the ELM

The physical basis for the ELM was given in equation 3.12 (Schott 1997) and is repeated below:

$$L' = L_u + (E'_s + L_d)T_r\pi^{-1}\rho$$

Note that the above formulation suppresses dependence on wavelength. This relationship will not be constant across all wavelengths or wavebands of a broad band sensor. Hence, the models below will need to be recalibrated for each waveband of the sensor of interest. The same formulation is given by Ferrier & Wadge (1996), although they omit the L_d term in equation 3.12.

This representation of the system is convenient, since the equation gives the relationship between at–sensor radiance and at–surface reflectance in a linear form. This can be re–written as:

$$L = a + b\rho \tag{3.14}$$

where:

$a = L_u$;

$b = (E'_s + L_d)\pi^{-1}T_r$

Schott (1997) proposes that a and b are derived using regression and that equation 3.14 is then rearranged for reflectance. This approach is problematic. Regression is based on minimising the variance in the predictor variable and assumes that the dependent variable is known *without error* (Neter et al. 1996, Webster 1989). Hence, rearranging a regression equation to predict the independent variable is controversial (see discussion at the end of section 2.1.4). Therefore, the following approach is adopted here. The remote sensing system gives an image of at–sensor radiance, hence L is known at all locations in the image. Since the objective is to predict ρ, equation 3.14 is rearranged as follows:

$$\rho = \frac{L-a}{b} = \frac{L}{b} - \frac{a}{b} = a' + b'L \tag{3.15}$$

Rearranging equation 3.14 invokes *no* assumptions about the regression model. It is also interesting to note that this provides a *mathematical* explanation of why the y-intercept term is negative (see also figure 3.5). The physical interpretation of this is that, at the x-intercept where $\rho = 0$, $L = L_u$.

If ρ can be determined at a series of representative locations throughout the image then this can be used as the basis for a regression model. In adopting the regression model, it

is assumed that the X variable is known without error. However, L is not, strictly, known without error, although it is sufficient to take the recorded value of L as a marker. The true value, denoted L^*, can then be thought of as a random variable, where $L = L^* + \delta$, and where δ is an error term. The fact that L^* is a random variable does not matter, since it is L and *not* L^* that is used in the regression model (see section 2.1.4 for discussion). The regression model is then invoked as:

$$\rho_i = \beta_0 + \beta_1 L_i + \varepsilon_i \tag{3.16}$$

where ε is an error term and i refers to a specific measurement. The error term incorporates measurement error (in ρ), model uncertainty and natural variation which is not accounted for by the model. Natural variation might include spatial variation in reflectance and atmospheric variation. Model uncertainty includes uncertainty introduced by pairing reflectance and radiance measurements. Examination of spatial variation and methodologies for pairing reflectance and radiance measurements form important components of this thesis.

3.2.3.2 Assumptions of the ELM

It is clear that the ELM, as described in sections 3.2.3 and 3.2.3.1, imposes several *requirements* if it is to be successfully implemented. These requirements are often manifested as *assumptions*, where the user is not certain whether the requirements have been met. Furthermore, several decisions need to be made regarding the way in which the methodology is implemented. This section reviews and critiques the requirements, assumptions and methodological issues.

The first requirement is that the atmospheric and illumination conditions are spatially uniform over the image. This is because, in the estimation and prediction phases it is assumed that T_i, T_r and L_u (see equation 3.12) are constant over the entire image. This requirement is difficult to test, although it can be mitigated by conducting the airborne survey in stable atmospheric conditions and seeking field sites that are close to the study areas (Kaufman 1989). Nevertheless, atmospheric conditions are known to change over short time intervals and distances (Milton & Goetz 1997). Similarly, the path length will vary most for wide field-of-view sensors. In areas of variable terrain atmospheric and illumination conditions and path length may all vary (Schott 1997).

The second requirement is that the atmospheric conditions and viewing and illumination conditions should not change between the over flight and the time that the field spectra are measured. Strictly, this is the case, since reflectance is a function of surface properties, viewing and illumination geometry and atmospheric conditions (see section 3.1). However, in terms of characterising the *apparent* reflectance of the GTs, it is argued here that this is an assumption about the GTs, rather than atmosphere or viewing and illumination conditions. Hence it is assumed that GTs are insensitive to viewing and illu-

mination conditions and are temporally stable. Indeed, this requirement lies behind the establishment of *pseudoinvariant features* that are used for atmospheric correction (Schott et al. 1988, Anderson & Milton 2006).

Further requirements are made about the properties of the GTs, the reflectance of which is measured in the field or laboratory (Smith & Milton 1999, Clark et al. 2000). These are listed and then discussed in detail below.

1. the GTs are clearly identifiable and accessible in the image and in the field;
2. the GTs are spatially extensive and substantially larger than the pixel size;
3. there are no adjacent obscurations which may limit illumination of the surface;
4. the reflectance of the GTs should bracket the range of reflectance values found in the image;
5. the GTs are spatially homogeneous in the spectral domain;
6. the GTs are spectrally "bland", *i.e.*, they have similar reflectance at all wavelengths;
7. the targets remain spectrally stable over time;
8. the GTs are flat and all at the same altitude;
9. the GTs exhibit Lambertian properties over all relevant viewing and illumination angles.

First, it is essential that the GTs can be identified in the image and on the ground and are accessible, so samples can be taken. This is so that the user can match up the field measurements with the spatially coincident remotely sensed measurements. This is a straightforward, but important point. However, few authors explicitly report formally recording the location of the GTs (Ben-Dor et al. 1994, Ferrier 1995, Ben-Dor & Levin 2000, Karpouzli & Malthus 2003). Several authors *imply* that the target which is sampled can be located in the image, but do not state that the location of the target is formally recorded (Kruse et al. 1990, Farrand et al. 1994, van der Meer 1994, Dwyer et al. 1995, Ferrier & Wadge 1996, Smith & Milton 1999, Perry et al. 2000, Yang et al. 2000). Furthermore, *none* of these authors report explicitly and formally recording the *location* of each field *sample*. The impact of this is discussed in detail below.

The requirement that the GTs are spatially extensive and substantially larger than the pixel size is made to reduce the possibility of adjacency effects (Smith & Milton 1999). This may occur because the point spread function of the airborne sensor is typically larger than the pixel size (Smith & Milton 1999) or because radiance upwelling from adjacent pixels is scattered into the target–sensor path (Kaufman 1989, Schott 1997) or because there are mixed pixels. These effects will be most significant where two or more surfaces with very different reflectance properties lie adjacent to each other (Kaufman 1989). Smith

& Milton (1999) recommend that GTs should be at least three times the size of the pixel (their investigation uses airborne data with a pixel size of the order of 1 to 10 m). Kaufman (1989) demonstrated that, for satellite data, adjacency effects may be significant at distances up to 2000 m, especially when the optical thickness is high. However, for airborne sensors Gu et al. (1992) did not find evidence of adjacency effects for images flown at less than 2500 m above ground surface. Nevertheless, if adequately sized GTs are not available, this will limit the possibility of applying this technique to a given image. This requirement may also be made to allow averaging over several pixels in order to reduce noise in the remotely sensed data (Clark et al. 2000).

The third requirement, that there should be no adjacent obscuration that may limit the illumination of the surface, ensures that there is consistency between airborne and field measurements and between field measurements taken at different times. This is because there will be no shadow cast on the surfaces that are being sampled (Schott 1997). As before, if no appropriate GTs are available, this will limit the possibility of applying this technique to a given image.

The fourth requirement is that the reflectance of the GTs should bracket the range of reflectance values found in the image. This is imposed so that it is not necessary to extrapolate to predict the value of reflectance at every location in the image (Smith & Milton 1999). If this requirement cannot be met, it may be proposed that the GTs should bracket the range of reflectances of surfaces that are of interest to the investigator. If no appropriate GTs are available, this will lead to an increase in the uncertainty of prediction where extrapolation takes place.

The requirement for GTs to be spatially homogeneous in the spectral domain is given by many authors (Kruse et al. 1990, Ben-Dor et al. 1994, Farrand et al. 1994, Dwyer et al. 1995, Ferrier & Wadge 1996, Perry et al. 2000, Karpouzli & Malthus 2003). In principle, this should simplify sampling of the target and reduces the possibility of including artifacts which may exist in surfaces of mixed composition (Clark et al. 2000). Furthermore, where targets are of mixed composition, failure to sample the various materials proportionally can lead to inaccurate aggregate averages of reflectance (Milton et al. 1997). This is discussed further in section 3.2.3.3. Where laboratory spectra are gained from field samples this requirement is extended to say that the *composition* of the targets should be uniform (Kruse et al. 1990, Ben-Dor et al. 1994, Farrand et al. 1994). The authors tend to *assume* that the target is homogeneous (perhaps based on visual examination) rather than testing that this requirement is met. However, it has been shown that surfaces that appear to be spatially uniform may exhibit spatial structure and this may vary with wavelength (Atkinson & Emery 1999). There are occasional examples where the GTs have been examined and found not to be homogeneous (Ferrier 1995, Salvatori et al. 2000). In particular,—Salvatori et al. (2000) used geostatistics to demonstrate the existence of spatial structure in concrete and asphalt GTs.

The requirement for GTs to be spectrally bland is made because it reduces the possi-

bility of artifacts caused by inaccurate spectral sampling (Smith & Milton 1999, Clark et al. 2000). If the targets have strong absorption or reflectance features then inaccurate characterisation of these can lead to artifacts in the atmospheric correction. This issue may be a particular problem for hyperspectral data and is less important for multispectral data. Furthermore, it may be noted that it may not be possible to find a sufficient number of spectrally bland targets such that spectrally variable targets, such as grass, have to be used (Smith & Milton 1999). In such situations, it is worthwhile to check for artifacts when using hyperspectral data. Finally, note that artifacts due to absorption and reflectance features may also arise with other atmospheric correction approaches (Roberts et al. 1986, Perry et al. 2000, Clark et al. 2000).

The requirement for GTs to remain spectrally stable over time is to reduce the impact of spectral instability where it is not possible to obtain field measurements at the same time as the flight or satellite overpass (Smith & Milton 1999, Milton et al. 2000, Anderson & Milton 2006). Indeed, this is generally not possible for logistic and operational reasons. Hence, the user needs to be confident that the reflectance will not change in the interim period which may last between a few hours and several days. Indeed, where targets remain stable over long time periods, the need to measure them for each flight or over-pass may be eliminated. This is the principle behind the use of pseudo–invariant features (Schott et al. 1988, Milton et al. 2000, Teillet et al. 2006).

The requirement for the GTs to be at identical altitudes is set so that the path–length will not vary between GTs (Schott 1997). Ferrier (1995) notes that non–adherence to this requirement may have introduced errors into their implementation of the ELM, although they do not quantify how. If the surfaces are not level then this will lead to obscuring of incident direct and diffuse radiation. This can be corrected for if the surfaces are Lambertian (Schott 1997).

Finally, the requirement that the targets are Lambertian is important since the reflectance of the targets may not be sampled under the same viewing and illumination geometry as the remotely sensed data. Furthermore, the viewing and illumination geometry may change between subsequent reflectance measurements. Specifying Lambertian surfaces limits any change in reflectance that occurs due to changes in viewing and illumination geometry or illumination conditions. It should be noted that apparently Lambertian surfaces may exhibit non–Lambertian properties at extreme illumination geometries (Rollin et al. 2000). Hence, it is advisable to select surfaces that are "near" Lambertian (Smith & Milton 1999) and not to make measurements at extreme solar zenith angles or under very diffuse illumination conditions. Some authors have posed exceptions to this requirement. Farrand et al. (1994) measure reflectance under laboratory conditions and specifically control for viewing and illumination conditions. Moran et al. (2001) and Anderson & Milton (2005) adopt the model of Walthall et al. (1985), which is a simple empirical BRDF model which models reflectance as a function of the solar zenith angle (SZA). They find that this is sufficient, providing that the angular reflectance properties vary in a consistent and predictable way.

The requirement for Lambertian targets is also important since reflectance is a function of surface and atmospheric properties and viewing and illumination conditions. If the atmospheric conditions change between the time of the flight and the time of the field measurements, then near–Lambertian targets will be affected less than non–Lambertian targets.

As with the assumption about homogeneous surfaces, it is generally *assumed* that the selected targets are Lambertian. However, it is rare for this assumption to be tested specifically.

The above discussion has covered the key requirements and assumptions that are made about the GTs and atmosphere. However, in any operational situation it may not be possible to meet or test these requirements in full. The following section discusses the implementation of the ELM.

3.2.3.3 Selection of ground targets

In the previous section, nine requirements, defining appropriate GTs, were set out. If requirements 2, 3, 4 and 8 are not met, then it might be concluded that the ELM may only be implemented with caution. However, it is proposed that measures can be taken in situations where requirements 1, 5, 6, 7 and 9 are not fully met.

When the GTs cannot be precisely located within the image Ferrier (1995) and Ferrier & Wadge (1996) report reduced accuracy in the ELM. This may be due to lack of precision in locating the GTs or the field samples or to imprecise geometric registration of the image. These effects are considered in detail in this work. Modern GPS and surveying technologies mean that it ought to be possible to record the location of the GTs (Karpouzli & Malthus 2003) and even individual measurements.

The issue of the homogeneity of the GTs requires careful consideration. It is proposed that it is unsatisfactory to assume that the targets are spatially homogeneous without testing this assumption. This is particularly important since it is known that surfaces that are *visually* homogeneous may exhibit spatial structure that varies with wavelength (Atkinson & Emery 1999, Salvatori et al. 2000). A detailed geostatistical examination may be conducted using field samples, providing the location of those samples is recorded. Where remotely sensed data from a previous campaign are available this assumption could be examined using those data. Such an examination is implemented in chapter 5.

To ensure homogeneity of the GTs, some authors have used GTs that are specifically constructed for the purpose (e.g., Maracci et al. 1981, Schott 1997, Perry et al. 2000). This approach is not always feasible or appropriate.

As discussed in the previous section, it may not be possible to find such surfaces that are spectrally bland. For example, bare rock or man-made surfaces (such as concrete and

asphalt) may have biological growth that is not clear to the naked eye (Clark et al. 2000, Anderson & Milton 2006). Indeed, grass is recommended as a candidate GT by Smith & Milton (1999) owing to its brightness in the near-infrared.

It should be noted that the surface may change between the time of the flight and the time that it is sampled. This may lead to change in the reflectance properties. This concern is, perhaps, most acute for laboratory samples, although Ben-Dor & Levin (2000) are careful to limit the disturbance and suggest that it is not a major problem in their study. Surfaces may change over time becauase of physical damage, vegetation growth, or because of wetting of the surface (Milton et al. 1996, Lawless et al. 1998, Smith & Milton 1999, Karpouzli & Malthus 2003), so it is important to consider these issues when selecting GTs. Lawless et al. (1998) report on a study of the spectral stability of asphalt and concrete targets. They conducted repeat measurements of reflectance of these two targets five times over a period of six weeks, using a broad-band Milton Multiband Radiometer (MMR) (Milton 1982, Milton 1987). However, Anderson & Milton (2006) found seasonal patterns in the reflectance of concrete on longer (annual) time scales which they attribute to the growth of biological material on the surface. Anderson & Milton (2006) also found an intra-day variation in reflectance of concrete of up to 2% in visible and NIR wavelengths which is correlated with an increase in relative humidity and the onset of a sea-breeze. They attribute this primarily to an increase in biological activity associated with the increase in humidity. Temporal variation in reflectance of GTs is recognised as an important issue, but lies outside the scope of this thesis.

Finally, the assumption that the GTs should exhibit Lambertian properties may be evaluated by reference to the literature on the angular reflectance properties of similar targets or by making measurements over the specific GT. Whilst the latter may be regarded as preferable, it demands much time and many resources and may not be feasible. Some analysis of bidirectional effects of GTs exists in the literature. Staenz & Itten (1982) find that asphalt and concrete targets generally exhibit Lambertian properties, although an exception is found for strongly grooved surfaces. Anderson & Milton (2006) find that, for nadir measurements, reflectance decreases as solar zenith angle increases – although this effect is largest when measurements are taken substantially before or after solar noon. Similarly, Clark et al. (2000) recommend that field measurements are taken as close to solar noon as possible. This area is given limited attention in this thesis.

3.2.3.4 Sampling reflectance of the ground targets

Given the discussion in the previous sections, it is clear that careful consideration needs to be given as to how the GTs are sampled.

First, it is useful to consider how reflectance is being measured. The airborne or satellite–borne instrument is likely to have a small instantaneous field of view. For example, the angular IFOV of the Airborne Thematic Mapper sensor used in this project is 0.14^o (2.5

mrad). The NERC ITRES CASI-2 instrument has an IFOV of 0.115^o. Spectrometers, such as those loaned by the NERC EPFS, typically have an IFOV of between 3^o and 10^o depending on the lens that is fitted. Hence, the angular sampling of the field or laboratory and air– or satellite–borne sensors are very different. This difference in the angular sampling is an inevitable consequence of sensor design and purpose and while cannot be addressed, is noted here for completeness.

Next, the viewing and illumination geometry need to be considered. Where field measurements of reflectance are used most authors do not report on the illumination or viewing geometry, although some report using a nadir viewing radiometer or spectrometer (e.g., Milton et al. 1997, Karpouzli & Malthus 2003) report using a nadir viewing radiometer. A nadir-viewing radiometer is used for the experimental work carried out for this thesis. Some authors recommend monitoring the atmospheric conditions and the diffuse to global (DG) ratio whilst undertaking field measurements (Milton et al. 2000, Clark et al. 2000, Anderson & Milton 2006) so that changes in illumination conditions can be established.

Some authors use laboratory, rather than field measurements of reflectance, allowing careful control of the material as well as the viewing and illumination conditions. Farrand et al. (1994) specifically controls for viewing and illumination conditions, although other authors do not report on this (e.g., Ben-Dor et al. 1994, Ben-Dor & Levin 2000). Whilst laboratory measurements are undoubtedly valuable when seeking to quantify and understand the reflectance properties of different surfaces, this approach is questionable for two reasons. First, laboratory conditions cannot fully recreate atmospheric illumination (Milton 1987, Milton et al. 1995). Second, the laboratory samples may not accurately represent the field environment. Laboratory samples are not used in this thesis.

The reflectance of GTs may change over seasonal and diurnal time-scales. Obvious changes may be due to wetting and drying or to substantial vegetation growth. However, changes may also occur due to biological activity that may not be obvious to the naked eye (Anderson & Milton 2006). This is an active and important research area (Milton et al. 1996, Anderson & Milton 2005, Anderson & Milton 2006) that has previously received limited attention. It poses substantial challenges both for quantifying these changes and for modelling changes in reflectance. It is important for the selection and use of GTs both for the ELM and for vicarious calibration. The importance of this issue is recognised and, during fieldwork effort was taking to mitigate these effects by conducting fieldwork as close to solar noon as possible and by selecting stable atmospheric conditions. However, detailed analysis lies outside the scope of this thesis.

When measuring the reflectance of the GTs using a spectrometer or radiometers the size of the areal IFOV will be much smaller than the ERE or pixel size measured by the remote sensor. For example, Salvatori et al. (2000) report using nadir viewing ASD FieldSpec FR™ fitted with a 8^o foreoptic mounted 3.6 m above the surface, yielding a 0.5 m diameter IFOV. Such information is typically not reported – although this example does

demonstrate that the support of the field measurement is likely to be very different from the support of the remotely sensed measurement (the pixel size). This has two important effects. First, measurements made on different supports are known to have different statistical properties (Blalock 1964, Heuvelink & Pebesma 1999, Bierkens et al. 2000). The validity of combining measurements that are defined on different supports within a model without specifically addressing this issue has recently been questioned (Heuvelink & Pebesma 1999, Dungan 2002). Typically, the measurements are averaged over the GT, giving a per GT measure of reflectance. This issue receives substantial attention in this thesis. Second, it may not be possible to find surfaces that are comprised of only one cover type and the GT may have features which have different reflectance properties (such as road–markings, or vegetation). These may make a significant contribution to the reflectance of a pixel sized area (Milton et al. 1997). In such circumstances, this effect should be accounted for, either by (i) measuring reflectance over pixel sized supports (this may involve placing the radiometer or spectrometer on the top of a tall mast); (ii) intensively sampling within a pixel sized area; or (iii) sampling the reflectance of the different components and computing the weighted-average reflectance, where the weights are proportional to the area covered by each component.

Finally, care needs to be given to determining both the number of field measurements that are required and the sampling regime. Smith & Milton (1999) take between 5 and 35 measurement for each GT (approximately 9 m by 9 m in area) and Karpouzli & Malthus (2003) take between 15 and 60 measurements per target. Farrand et al. (1994) report measuring 100 spectra for one of their GTs, but do not report how many are taken over the second GT. These authors do not typically specify a sampling scheme. Clark et al. (2000) do not take point-support measurements but configure the spectrometer to average 60 0.1-second integrations over a 6 second period. The operator then traverses the site and the resultant measurement yields a spatial average over 4 to 8 m (depending on the speed the operator walks. This approach is used to sample large areas quickly. However, with such an approach it is not possible to give a location to individual measurements. It is clearly important to take enough measurements to characterise each surface, although there is no indication of what constitutes "enough" measurements. Furthermore, to account for spatial structure in the reflectance of each GT it is necessary to sample at a range of lags (Salvatori et al. 2000) and to record the location of each measurement. Furthermore, providing a full characterisation of GTs can be substantially time consuming. Clark et al. (2000) note that to achieve accurate atmospheric correction approximately one person-month was spent characterising the GTs. However, many other researchers spend substantially less time and effort characterising GTs.

3.2.3.5 Implementation of the ELM

As introduced in section 3.2.3 the ELM is based on a linear relationship between reflectance and radiance. To obtain the slope and intercept parameters of the linear model

it is necessary to establish radiance–reflectance data pairs that can be used to estimate the model parameters. At least two such data are required, in which case a line is simply drawn through them and the slope and intercept parameters calculated. If more than two data pairs are available the parameters are estimated using regression modelling, as discussed in section 2.1. When regression is used the model allows for assessment of uncertainty in estimation and prediction.

A major issue here is the approach that is used for determining these data pairs. The approach that is typically adopted in the literature is to identify the GT in the image and to calculate the mean radiance (or DN) by averaging all the pixels in the GT. Field measurements of reflectance are then obtained and averaged to yield the mean reflectance value for the GT. In this thesis, this approach is referred to as the *typical approach.*

Inspection of the reflectance data can aid the assessment of whether the GT is suitable for use in the ELM. Inspection of the histograms of reflectance measurements can give insight here. Outliers may be indicative of problems in the measurement system or of genuine bright or dark patches in the target. A bi-modal or multi-modal distribution may be indicative of a target of mixed composition. If the location of each measurement is known then the spatial structure in the data can be examined. If the time when each measurement is taken is known then the reflectance data can be examined for evdicence of directional effects.

The above mentioned "typical approach" is used almost universally when implementing the ELM. However, it is problematic for various reasons, as follows:

1. The support of the GTs are ambiguously defined. It is not made clear what size they are or where the boundaries lie. Indeed the support may vary between GTs – or even between the image-based and field-based definitions of the same GT. This is problematic for two reasons:

   (a) It is not possible to state clearly the support of the data for estimation and prediction.

   (b) The data pairs obtained using this approach are typically used to estimate the parameters of the ELM. These are then applied to the remainder of the image. This approach is problematic unless the objective is to predict on GT sized supports (which is rarely the case). If the objective is to predict on pixel sized supports the approach is problematic, as discussed in section 2.2.1, since the supports used for estimation and prediction will be different. The parameter estimates and predictions based on these estimates may be different from those based on estimates obtained from pixel-sized targets. For spatially homogeneous targets and the linear model, no significant change in the estimates of the slope and intercept parameters is expected (Heuvelink & Pebesma 1999) – although the uncertainty estimates will differ. For inhomogeneous targets

MAUP effects may lead to different parameter estimates for pixel-sized and GT-sized supports.

2. Averaging over GTs reduces the information content of the data. As mentioned above, this is problematic if the objective is to predict on pixel sized supports. This is because averaging reduces the variability and may lead to unrealistically low values of the variance and, hence, unrealistically tight confidence and prediction intervals. However, when the number of GTs used in the regression is small this effect may be dramatically reversed. This is because the t–values used in equations 2.12 and 2.13 are dependent on the degrees of freedom (ν) and increases sharply when ν is small. This is particularly clear when $\nu \leq 5$ but persists until $\nu \approx 20$[2]. This sharp increase reflects the uncertainty in the parameter estimation owing to the small sample size and may work against the decrease in the variance caused by averaging. Note that $\nu = n - 2$ for the simple linear model adopted in equation 3.16 (n is the number of data). The ELM is generally implemented with a small number of GTs and the nine GTs used by Karpouzli & Malthus (2003) is unusually large. Hence, for a more usual number of GTs (e.g., 3 or 4) the confidence intervals on the regression parameters may be wide and even lead to rejection of the regression model. The hypothesis tests on the regression parameters have *never* been reported for the ELM.

3. When averaging over the GTs it is important to consider the field sample size and configuration necessary to yield a stable estimate for the mean. This is often not stated and it may range from less than ten measurements per target (e.g., Roberts et al. 1986) to several hundred (e.g., Gu et al. 1992, Clark et al. 2000). Hence there is a need for a quantitative assessment of the required sample size.

These above issues are exacerbated where it is not possible to find spatially homogeneous targets. In such cases, it may be necessary to proceed using sub-optimal targets.

These above issues are given detailed empirical attention in this thesis and form the bulk of the material presented in chapter 5. However, implementing the ELM on a per-pixel basis introduces two further important considerations:

1. Positional accuracy of the field-measurements, both relative to themselves and relative to the image

2. Limitations posed by the regression assumptions. Notably, the regression residuals tend to be heteroskedastic and spatially autocorrelated.

These issues are addressed in chapters 6 and 7 respectively.

This thesis is, therefore, concerned primarily with the spatial sampling and model implementation issues surrounding the ELM. It is recognised that the directional and temporal

[2]Tables showing the percentiles of the t–distribution given in many text books (e.g., Neter et al. 1996).

stability of the GTs is of major importance. However, these issues lie beyond the scope of this thesis and are given only limited attention.

3.3 Summary

This chapter has provided an overview of the physical basis of remote sensing. It has then gone on to review methods for atmospheric correction of airborne and satellite optical data. In particular, the Empirical Line Method (ELM) received detailed attention.

Chapter 4

Field site and data procurement

This chapter introduces the field site and instrumentation used. It then goes on to introduce the fieldwork objectives and experimental design.

4.1 Description of field site: Thorney Island

Thorney Island (GR: 50^o 48' 35" N, 0^o 55' 19" W) is located in Chichester Harbour, West Sussex on the south coast of the British mainland. The "island" has been joined to the mainland for approximately 125 years, following reclamation of tidal mud flats in the 1970s. The site is owned by the UK Ministry of Defence (MOD). It was used by the Royal Air Force (RAF) from 1935 up until 1984, when it was taken over by the Army. Access to the site is controlled by the MOD hence, although there is some access to hobbyists and other scientists and engineers, much of the site is largely disused. This generally allows undisturbed access for scientific purposes. The site has been used by researchers in the School of Geography and NERC Equipment Pool for Field Spectroscopy (EPFS) (as was) over the last decade (Milton et al. 1996, Lawless et al. 1998, Smith & Milton 1999, Salvatori et al. 2000, Anderson 2005, Anderson & Milton 2005, Anderson & Milton 2006).

Thorney Island's status as former air field means that it has a runway comprised of asphalt and concrete surfaces. These range between 50 m and 100 m in width and hence are of suitable spatial extent for use with fine spatial resolution satellite (e.g., IKONOS, QuickBird) and airborne (e.g., ATM, CASI) sensors. The grass surrounding the runway is also kept cropped and has been used as a GT by Smith & Milton (1999). Hence, these three surfaces provide potential GTs with a range of reflectance values at optical wavelengths. For the research conducted in this thesis, fieldwork focused on the asphalt, concrete and grass surfaces at the southern end of the runway, the approximate location of which are shown in figure 4.1.

Thorney Island exhibits only slight relief, meaning that all potential GTs are at approximately the same attitude. The lack of relief also simplified geometric correction of the

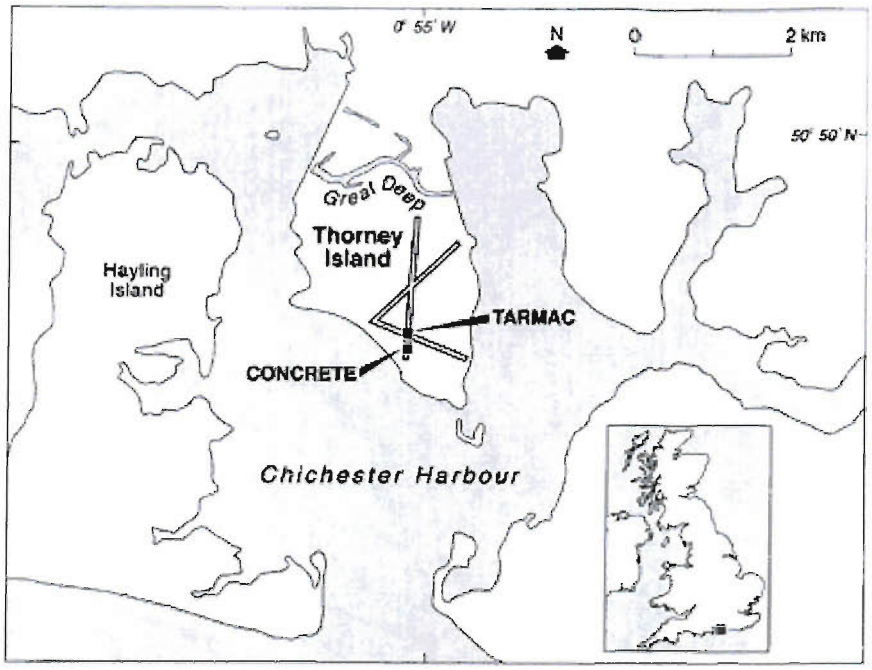


FIGURE 4.1: Map of Chichester Harbour, including Thorney Island. Approximate locations of the concrete and asphalt (tarmac) GTs are shown. The grass GT is slightly to the east of the asphalt GT.

image, as discussed in section 4.2.5. Furthermore, the runway is characteristic of other WWII runways (e.g., Chilbolton, Hampshire), which might provide suitable sites for atmospheric correction and vicarious calibration elsewhere in the UK. Finally, Thorney Island is easily accessible from Southampton, thus simplifying the fieldwork logistics.

4.2 Airborne imagery: NERC ARSF

The UK NERC provides an Airborne Remote Sensing Facility (ARSF), primarily for use by the UK academic community. Groups of researchers can make a bid to have their sites over flown and the process is similar to that undergone when applying for a research grant. The application is graded according to its scientific merit and the team then liaise with the flight crew to establish the logistics of the mission. Some published information on the facility is available (Roy et al. 1997, Wilson 1997) however most technical and practical information about the facility is primarily available in the "grey" literature in the form of user manuals and documentation available from the ARSF via the Internet[1]. Further useful information is often not formally documented and may be dispensed at local science meetings (such as the annual workshop organised by the ARSF and the meetings of the UK Remote Sensing and Photogrammetry Society) or via personal communications.

The ARSF operates a Dornier 228–101 D–CALM aircraft with an array of instruments mounted on it. The remote sensing instruments available are the Daedelus 1268 Airborne

[1]For further information on the NERC ARSF see: `http://arsf.nerc.ac.uk/`

Product	Level
Level 0	Raw "sensor format" data at original resolution.
Level 1a	Level 0 data reformatted to image files with ancillary files appended.
Level 1b	Level 1a data to which radiometric calibration algorithms have been applied, to produce radiance or irradiance, and to which location and navigational information has been appended.
Level 2	Geophysical or environmental parameters derived from Level 1a or 1b data, may include atmospheric correction.
Level 3a	Level 1b or 2 data mapped to a geographic co–ordinate system using on–board attitude and positional information only.
Level 3b	Level 1b or 2 data mapped to a geographic co–ordinate system using on–board attitude and positional information with additional ground control points.
Level 4	Multi–temporal/multi sensor gridded data products.

TABLE 4.1: Table showing definition of the HDF product levels. Source: NERC ARSF (`http://arsf.nerc.ac.uk/data/`).

Thematic Mapper (ATM), the Itres Instruments Compact Airborne Imaging Spectrometer 2 (CASI) and the Wild RC camera. The CASI is complemented by the Incident Light Source (ILS) instrument which measures downwelling radiation at the aircraft. The aircraft is fitted with an array of Global Positioning System (GPS) instruments and an inertial navigation system. This allows the position and trajectory (including variation in pitch, roll and yaw) of the aircraft to be recorded. This information is used for geocorrection and geocoding of the remotely sensed data.

Together the instruments form the Integrated Data System (IDS) which has been developed to integrate the imagery with navigation and altitude data provided by onboard Global Positioning Systems (GPS) receiver. The data can subsequently be processed to yield a georeferenced digital data set *without* the need for ground control points, giving a level 3a product.

ARSF data are supplied as a level 1b hierarchical data format (HDF) product following the system defined by NASA[2] (table 4.1). ARSF provides software that allows the data to be processed to a level 3a or 3b product. Further processing of the data requires that the user defines their own methodology and processing technique. Definition of the various levels of data product are given in table 4.1.

[2]For more information on HDF can be found from the NERC ARSF (`http://arsf.nerc.ac.uk/data/`) and from NASA (`http://eosweb.larc.nasa.gov/HBDOCS/hdf.html`).

IFOV	2.5 mrad
Samples per scanline	938
Digitised FOV	$\pi/2$ rad
Scan Rate	12.5, 25 or 50 Hz
Radiometric Resolution	16 bit
Temperature Reference	Two black–bodies for calibration of thermal channel

TABLE 4.2: AZ–16(ATM) specification. Source: NERC ARSF (`http://arsf.nerc.ac.uk/instruments/atm.asp`.

ATM band	Wavelength μm	Landsat TM band	Wavelength μm
1	0.42 – 0.45	-	-
2	0.45 – 0.52	1	0.45 – 0.52
3	0.52 – 0.60	2	0.52 – 0.60
4	0.60 – 0.62	-	-
5	0.63 – 0.69	3	0.63 – 0.69
6	0.69 – 0.75	-	-
7	0.76 – 0.90	4	0.76 – 0.90
8	0.91 – 1.05	-	-
9	1.55 – 1.75	5	1.55 – 1.75
10	2.08 – 2.35	7	2.08 – 2.35
11	8.50 – 13.0	6	10.4 – 12.5

TABLE 4.3: Table showing ATM and Landsat TM wave bands. Source: NERC ARSF `http://arsf.nerc.ac.uk/instruments/atm.asp`) and Campbell (1996).

4.2.1 The Airborne Thematic Mapper (ATM)

The Airborne Thematic Mapper (ATM) instrument provided by the ARSF is based on the Daedalus AADS1268 instrument. This system was overhauled and upgraded during the 1990s to meet the AZ–16 specification (table 4.2). The ATM is a passive response sensor that quantifies the radiation upwelling from the Earth's surface in 11 spectral bands (table 4.3). The ATM wavebands cover the visible and the near-infrared (bands 1 to 8), the shortwave infrared (bands 9 and 10) and the thermal infrared (band 11). Several of these bands have been set to simulate the seven spectral bands of the Landsat Thematic Mapper (TM) (Wilson 1997).

The ATM is a whisk-broom type instrument (Campbell 1996). The scan mirror has three synchronised speeds (12.5, 25 and 50 Hz), which allows the scan rate to be adjusted as appropriate for a given flight speed and altitude in order to ensure that there is not too much over- or under-sampling in the along track direction. The sensor has a fixed instantaneous field of view (IFOV) of 2.5 mrad ($\sim 0.14°$), so the effective resolution element (ERE) of the sensor is dependant on the altitude and viewing angle. For each scanline 938 samples are taken and the digitised field of view (FOV) is $\pi/2$ (90°). This gives an angular sampling interval of $\frac{\pi/2}{938} = \pi/1876$ radians ($\sim 0.096°$), which may lead to

IFOV: Across track	54.4^o
IFOV: Along track	0.1151^o
Spectral range	405 - 950 nm
Spatial samples	512 spatial pixels (across track)
Spectral samples	288 at 1.8 nm intervals
Dynamic range	12 bit

TABLE 4.4: NERC CASI specification. Source: NERC ARSF (`http://arsf.nerc.ac.uk/documents/casi2.pdf`.

over sampling in the across track direction. A general specification is given in table 4.2.

4.2.2 Compact Airborne Spectrographic Imager (CASI)

The CASI provided by the ARSF is the Itres[3] CASI-2 instrument[4,5]. It contains a two-dimensional CCD array-based pushbroom imaging spectrometer, described in more detail in table 4.4. A full 512 pixel hyperspectral (288 bands) image cube can be obtained (full-frame). However, in practice the integration time is so high (1–2 s), that this is only feasible if the instrument is mounted on a tripod rather than an aircraft. Hence, for CASI flights there are two basic modes of operation which represent a trade-off between spectral and spatial resolution. In spatial mode up to 18 programmable spectral bands can be obtained for the full 512 pixel swath. Pixel resolutions vary from sub-metre to 10 metres, dependent on the flying height. In spectral mode, the full 288 bands can be obtained but for more limited spatial sampling. For example, in enhanced spectral mode, a block of 101 adjacent pixels are recorded.

The ARSF CASI is offered with a default 12 band vegetation or ocean-colour band set, although the configuration can be specified by the user if required. The vegetation band set, used in this thesis is shown in figure 4.5.

4.2.3 Geometric correction of airborne imagery

Users of airborne remotely sensed data are familiar with images that are defined using a grid of, typically square, pixels. However, the raw data require considerable processing to get to this stage.

First, the pixel does not define the area on the ground that is imaged by the sensor. That area is known as the effective resolution element (ERE) and is typically larger than the basic pixel size. The proportion of the ERE that is represented by the pixel depends on the point spread function (PSF) of the sensor, the sensor configuration and the pixel size to which the data is processed (Richards & Jia 1999).

[3] `http://www.itres.com/`
[4] `http://arsf.nerc.ac.uk/instruments/casi.asp`
[5] `http://arsf.nerc.ac.uk/documents/casi2.pdf`

Channel	Centre (nm)	Width (nm)	Start (nm)	End (nm)	Purpose
1	450	20	441.53	459.17	Vegetation response (blue)
2	490	20	480.37	499.84	Vegetation response
3	552	10	547.74	556.63	Vegetation response (green)
4	670	10	665.57	674.74	Vegetation absorption maximum
5	700	10	694.28	703.27	Red-edge
6	710	10	705.07	711.06	Red-edge
7	740	10	735.66	744.67	Red-edge
8	750	7	746.47	753.68	Red-edge
9	762	5	760.9	764.51	Oxygen absorption
10	780	10	775.34	784.37	Vegetation reflectance maximum
11	820	10	815.13	824.18	Water absorption
12	865	10	860.46	869.54	NIR plateau

TABLE 4.5: Table showing the NERC ARSF CASI default vegetation wavebands. Source: NERC ARSF `http://arsf.nerc.ac.uk/documents/casi2.pdf`).

Next, the image is subjected to several forms of geometric distortion (Mather 1999, Richards & Jia 1999). At the edges of the image swath, the area imaged on the ground will be larger than that imaged at nadir. Since the sensor records at fixed angular increments, the spacing of the nominal pixel centres increases towards the edges of the image swath. Furthermore, the juxtaposition of the forward motion of the aircraft and the across track motion of the sensor leads to an "S–bend" effect in ATM imagery.. These are termed panoramic distortions.

ATM data are subject to further distortions, induced by the whiskbroom system. The scanner moves as the aircraft moves forward, leading to skewing of the scan lines in the along track direction (Richards & Jia 1999). Furthermore, as noted in section 4.2.1, the sensor generally over samples the image in the along and across track directions.

Finally, variation in altitude and in pitch, roll and yaw lead to distortions for both whiskbroom and pushbroom systems.

For satellite based sensors, distortions are induced by the Earth's curvature and rotation. However, these are not generally considered to be important for airborne data of the type used in this study (Mather 1999, Richards & Jia 1999).

By taking account of the above distortions it is possible to define the image on a grid. For many applications, it is necessary to geocode the image to a map co–ordinate system. In this situation, further distortions may be introduced by the topography of the landscape.

4.2.4 Preprocessing of the imagery

NERC ARSF provide data as a HDF level 1b product (see table 4.1). This was processed to yield a geometrically corrected HDF level 3a product. The HDF file can then be further processed to yield a binary file suitable for input into image processing software (such as Research Systems' ENVI) or for processing using custom built programs (e.g., C or IDL programs). To process the data from the level 1b product to level 3a, ARSF provide the `azgcorr` software. Further processing is conducted using `azexhdf`. `azgcorr` version 4.3.3 and `azexhdf` version 2.0.0, the latest available at the time of processing, were used.

Several options may be provided to `azgcorr` and `azexhdf`. Those options that actually alter the data content of the processed image are listed and discussed below. This is not intended to be an exhaustive list, since the specific configuration of options will vary according to the intended use. However, it does discuss those that affect the data processing conducted for this project. Options which alter the data format (rather than content) are not discussed. In particular, `azexhdf` is only used to convert data from HDF to some other user specified format (in this case the binary interleaved by line (BIL) format was used). The stated options are:

1. Geodetic datum shift (vertical geodetic transformation);
2. Map projection information (horizontal geodetic transformation);
3. Incorporation of a Digital Elevation Model (DEM).
4. Pixel size on the output image;
5. Resampling the output image;

As mentioned in section 4.2.1, the position of the aircraft is determined using a network of GPS. This position is given in three dimensions, relative to the European Terrestrial Reference System 1989 (ETRS89) (Ordnance Survey 1999). ETRS89 is defined by the World Geodetic System 1984 (WGS84) datum at 1989.0 (i.e. 00:00 hours on 1st January 1989). ETRS89 is realised by the European Terrestrial Reference Frame 1989 (ETRF89). To convert from the GPS co–ordinate system to a map co–ordinate system it is necessary to apply a geodetic transformation (Robinson et al. 1995). In the UK, the Ordnance Survey provides the required algorithms, parameters and data files to convert from ETRF89 to the UK National Grid, where the horizontal position is defined according to OSGB36 and the vertical position is defined by the Ordnance Datum Newlyn (ODN).

Within `azgcorr` the default (UK95) vertical and horizontal geodetic transformation use the 'Ordnance Survey recommended National Grid/ETRF89 Transformation Parameters 2 / 1995 version 1.2'. These use a grid contained within the software and has a stated accuracy of 2 m in the horizontal. This transformation has since been superseded (Colin Fane, Ordnance Survey, personal communication). The two transformations now provided

by the Ordnance Survey are the National Grid Transformation (OSTN97) to convert from ETRS89 to OSGB36 and the National Geoid model (OSGM91) to convert from ETRS89 to Ordnance Datum Newlyn (ODN). Hence, OSTN97 provides the horizontal transformation and OSGM91 provides the vertical transformation. OSTN97 has a stated RMS accuracy of 0.2 m and OSGM91 has a stated RMS accuracy of 0.1 m (95%) (Ordnance Survey 1999). These transformations may be applied by using the "UK99" option.

The Ordnance Survey have recently (late 2002) updated OSTN97 and OSGM91 to OSTN02 and OSGM02 respectively. In addition to providing a new standard with increased accuracy, these transformations now *define* the UK National Grid (Ordnance Survey 2002a). OSTN02 has a stated RMS accuracy of 0.1 m (1 standard deviation) and OSGM02 has a stated RMS accuracy of 0.02 m (2 standard deviations) (Ordnance Survey 2002a). These transformations were *not* available with `azgcorr` at the time of processing.

In order to gain an accurate geometric correction it is necessary to provide some information on elevation, since the ellipsoid (used for ETRS89) and the geoid (used for ODN) do not coincide. Hence, if elevation information is not provided, the software cannot calculate the distance between the scanner and the ground (the surface of ETRS89 is approximately 40 m below ODN sea level). If the user is confident that there is no relief in the landscape then it may be acceptable to provide a constant offset. However, if an accurate geometric correction is required, then it is necessary to use a digital elevation model (DEM). For example, a 1 m change in ground–sensor distance will lead to a 1 m displacement at the edge of the ATM swath (FOV = 90°. It was stated that `azgcorr` would accept either a digital terrain model (DTM) or contour data although, in practice, the software could only cope with a DTM.

When processing the level 1b product to level 3 it is necessary to specify a pixel size for the output image. Given the IFOV of the instrument and the flying height, it is possible to calculate a nominal pixel size at nadir. This changes where the scanner images the surface at off–nadir view angles and if the flying height or surface topography changes. Furthermore, as stated in section 4.2.1, the ATM *over samples* the image in the along– and across–track directions. Hence, the pixels in the output image *cannot* hold a direct relationship with the measurements that are actually made by the sensor. `azgcorr` provides three primary options for resampling the input image: a bi–cubic spline method, nearest neighbour interpolation and bilinear interpolation. It is commonly argued that using the nearest neighbour method is to be preferred, since this does not alter the input data (Mather 1999). However, at the most basic level, this is *not* correct, since `azgcorr` averages the input data before it resamples them (Andrew Wilson, NERC ARSF,personal communication). The bi–cubic spline method is the default option, since this aims to replicate the pixel value that would have been measured at any given map location. It also attempts explicitly to take account of the over sampling of the image.

4.2.5 Preprocessing the Thorney Island data

The Thorney Island data was processed using `azgcorr` and `azexhdf` as discussed above. For the vertical and geodetic shifts both the default (UK95) and upgraded (UK99) options were both used and the results explored. The Thorney Island site exhibits low relief, and is clearly close to sea level (a 5 m contour passes through the south western corner of the island). The need for a DEM was emphasised in section 4.2.4, so one was included for the preprocessing of the Thorney Island data.

Three DEMs were available for this work, as follows:

- The Ordnance Survey Land–Form PANORAMA®™ DEM;
- The Ordnance Survey Land–Form PROFILE®™ DEM;
- The LandMap DEM.

The PANORAMA® DEM is described in detail by Ordnance Survey (2001a). It is provided as 20 km by 20 km tiles and is available as contour data or as a DTM. The contour lines are digitised from the OS Landranger 1:50,000 scale map series where the contours are captured at 10 m vertical intervals. The contours have a stated global accuracy of ±3.0 m RMSE. This refers to the difference between the contour height recorded on the map and the measured height at the point on the ground represented by that contour. The DTM consists of height values at each intersection of a 50 m grid. The points are interpolated from the contour as well as tide line and spot height data using an unspecified interpolation technique. It is stated that DTM height accuracy is "*typically better than one half of the vertical accuracy of the source data* [i.e. 5 m]" (p 3.4), although the local accuracy is also dependent on the spacing and accuracy of the height data from which the DTM is interpolated. This is of particular concern since, according to the Landranger series, there are *no* contours on Thorney Island. The DTM of the whole of Thorney Island is at 3 m or below and the study area in the south of the island is at 2 m or below. This contradicts what is shown in the 1:25,000 OS Explorer map. That map shows a 5 m contour in the south–west of the island. Land–Form PANORAMA® data are available to the UK academic community for non–commercial research and study via the EDINA® Digimap® facility.

The PROFILE® DEM is described in detail by (Ordnance Survey 2001b). It is provided as 5 km by 5 km tiles and is available as contour data or as a DTM. The contour lines are digitised from the OS 1:10,000 scale map series where the contours are captured at 5 m vertical or 10 m vertical intervals. For Thorney Island and the surrounding area, contours were captured at 5 m vertical intervals. The contours were originally surveyed from an aerial photography campaign that began in the early 1960s and ended in 1987. The contours have a stated global accuracy of ±1.0 m RMSE. This refers to the difference between the contour height recorded on the map and the measured height at the point

on the ground represented by that contour. The DTM consists of height values at each intersection of a 10 m grid. The points are interpolated from the contour as well as tide line and spot height data using an unspecified interpolation technique. It is stated that DTM height accuracy is "*typically better than one half of the vertical accuracy of the source data* [i.e. 2.5 m]" (p 3.4), although the local accuracy is also dependent on the spacing and accuracy of the height data from which the DTM is interpolated. This is of concern on Thorney Island, since there is only one contour and the area is known to exhibit low relief. However, it is reasonable to assume that the PROFILE® contour and DTM data will be more accurate than the PANORAMA® data set, since it is a larger scale and more recent product. Ordnance Survey now dispense the PROFILE® DEM as their standard product.

The Landmap project was a joint project between University College London (UCL) and Manchester InforMation and Associated Services (MIMAS) (Kitmitto et al. 2000, Muller et al. 1999) . It has provided a DEM for the UK and Republic of Ireland together with a set of mosaiced remotely sensed imagery of the British Isles. The DEM was produced using Synthetic Aperture Radar (SAR) data from the European Space Agency's (ESA) ERS-1 and ERS-2 using interferometry (IfSAR). The practice of creating DEMs from SAR interferometry is widely documented (e.g., Rufino et al. 1998, Walker et al. 1999). The Landmap DEM has been created by combining four different ERS data sets (termed "[over]passes"), and consists of height values at each intersection of a 25 m grid. However, there is currently no accuracy assessment documented (either on the Internet or in the literature) for the 25 m DEM. Accuracy assessment was conducted on the "first pass" 30 m DEM (a pilot product that only used the ERS data from the first overpass) (Muller et al. 1999, Cross et al. 2000). That exercise compared the first–pass DEM to LiDAR data and to height data obtained from a field survey using kinematic GPS. That showed that the DEM was biased by between 1 and 70 m. Some early and unpublished accuracy assessment for the 25 m DEM showed a bias of 2.2 m and a standard deviation of 22.4 m (Kamie Kitmitto, MIMAS, personal communication). The Landmap DEM seems to be a promising product and could be useful for the processing of airborne data. This is because it is higher resolution (both spatially and vertically) than the PANORAMA® DEMs and is freely available (unlike the PROFILE® DEMs). However, the product is limited by the bias and high standard deviation shown by the accuracy assessment. Furthermore there is only a limited amount of documentation comprising short conference papers (Kitmitto et al. 2000, Muller et al. 1999) and a website (`http://www.landmap.ac.uk/`). There are few examples in the peer-reviewed literature which either documents the production or use of the data (Wilson & Atkinson 2005) and there are no technical manuals. In contrast the OS PANORAMA® and PROFILE® data sets have large user base and are produced by the national mapping agency. The objective of this project is *not* to conduct an exhaustive technical review of DEMs.

The PROFILE® DEM was chosen for use in this thesis because it is the finest resolution and most accurate DEM offered by the Ordnance Survey and the Landmap DEM was of

Date	24th July 2001
Time	08:56 – 08:59 GMT
Julian day	205
Flying height (GPS height)	2800 feet (853 m)
Ground speed	125 knots
Scan speed	50 Hz
Basic pixel size	2 m × 2 m
Flight direction	180° (North – South)
Flight line start	476440.1 108173.9
Flight line end	476110.4 100157.0

TABLE 4.6: Details of ATM flight over Thorney Island flown on 24th July 2001.

lower accuracy. Other DEMs are available in the UK. However, these are either of a lower resolution than Landmap or PROFILE® (Muller et al. 1999, Walker et al. 1999) or not available for the Thorney Island area.

In the analysis, a range of pixel sizes were used to explore the effects of different spatial resolutions of the remotely sensed data. For the ATM data set used in this project, the *base* pixel size chosen was 2 m. Further details about the ATM data set are given in table 4.6. The *GPS* flying height given in the flight log is 2800 feet (853 m). For Thorney Island, this approximates *above ground* to 800 m, since Thorney Island is approximately 5 m above sea level and the ETRS89 ellipsoid lies approximately 46 m ODN at Thorney Island. This approximate geoid-spheroid separation is found by querying the OSGM02/OSTN02 transformations via an OS-provided software tool (Grid InQuest v. 6.0.7)[6]. This allows the user to enter National Grid Eastings and Northings and ODN height and calculate latitude, longitude and ellipsoid height. By entering an ODN height of 0 m, the ellipsoid–geoid separation is found for a given point. This was conducted for 1200 points (on a 100 m grid) on Thorney Island, giving a mean geoid–ellipsoid separation of 45.58 m with a standard deviation of 0.02 m. The approximate 5 m ODN height of Thorney Island was based on examination of the OS 1:25,000 (Explorer™ 120) map. It is not possible to select a base pixel size that exactly represents the nadir pixel size for all scan lines so an approximately correct and convenient pixel size was chosen. The ATM has an IFOV of 2.5 mrad, giving a nadir pixel size of 2 m for a flying height of 800 m above the ground.

For the resampling, the bi–cubic spline method was chosen, for the reasons specified in section 4.2.4.

Independent data sets were used to assess the accuracy of the geometrically-corrected image that was output from `azgcorr`. At the first stage a visual comparison was made by overlaying OS Land–Line.Plus® data within the MapInfo Geographic Information System (GIS). Land–Line.Plus was the largest scale map data available via the EDINA® Digimap® facility and was also the product that ARSF benchmarked its geometric correction system against (Andrew Wilson, NERC ARSF, personnel communication). Land–

[6]Further information and the software tool can be found at: `http://www.gps.gov.uk`.

Line.Plus was available for Thorney Island at 1:2500 scale. Land–Line.Plus has a stated RMSE absolute accuracy of ±2.8 m (Ordnance Survey 2002b). This is a global statistic and may not strictly apply to the Thorney Island area.

Figure 4.2 shows the ATM image from 24th July 2001 with Land–Line.Plus data overlain. The main figure shows the main part of Thorney Island and the inset shows the primary study area. The pilot used the main runway (running from North to South) as a marker for the centre of the flight line. Visually, the geometric correction seems to be accurate at the centre of the swath, but is less accurate away from the centre. If accurate geometric correction had been required over the whole flight line then it would have been necessary to conduct further geometric correction using ground control points (GCPs). However, for this project, accurate geometric correction was only required at locations where field data were measured. These were all collected in the area shown in the inset and were deliberately placed close to the centre of the image swath. To test the *local* accuracy of the geometric correction GCPs were collected and compared to the location of these points in the image. GCPs were collected using two methods. First, points were surveyed on the ground, using a total station. This gave ten GCPs close to the locations where field measurements were taken. Further GCPs could not be collected by surveying owing to the lack of points that were clearly locatable both in the field and in the image. Further points along the runway were determined from the Land–Line data. It would have been preferable to determine the location of these points using surveying, since this would have been more consistent – however, this was not possible owing the limited line of site of the surveying instrument. Using the ten surveyed GCPs the RMSE was 0.84 pixels (1.67 m) with a maximum difference of 1.41 pixels (2.83 m). If the map–located GCPs collected along the runway were added, sub–pixel accuracy was also found. In contrast, GCPs located towards the edge of the image swath were over 10 pixels (20 m) in error. On the basis of this information it was decided that further empirical geometric correction was not required, since sub–pixel accuracy was found in the vicinity of the field measurements. Note also that additional geometric correction would require resampling which would alter the remotely sensed data still further.

Note that the geometric correction system is undergoing continual review (Andrew Wilson, NERC ARSF, personnel communication). Data obtained for a separate project in 2002 was also geometrically corrected and assessed visually. The geometric correction appeared to be much more accurate across the whole of the image swath than the 2001 correction. This will be encouraging news for future users of ARSF data.

4.3 Field spectroscopy

The term field spectroscopy, as the name suggests, refers specifically to the practise of taking spectral measurements of surfaces in the field environment. Spectral measurements are also taken in the laboratory environment. Indeed large blocks of concrete or sections of

FIGURE 4.2: Figure showing an ATM false colour composite (Red: Band 7; Green: Band 5, Blue: Band 2) image of Thorney Island (collected 24th July 2001). The insert shows the primary study area in greater detail. The image is overlain with OS Land–Line.Plus data. Validation GCPs are also shown. Validation GCPs are gathered either from the Land–Line.Plus map or via surveying.

grassland have been studied in the laboratory (e.g., Sandmeier et al. 1998). Such studies are valuable but pose two key limitations. First, the sample is disturbed, rather than being measured in situ. Second, it is not possible to fully recreate field illumination conditions in the laboratory (Milton 1987, Milton et al. 1995). There are, broadly, two types of instruments: those designed to collect data over a range of wavelengths (spectroradiometers) and those providing data for a few broad wavebands (multiband radiometers).

Field spectroscopy has a vital role in remote sensing research, serving various purposes (Milton 1987, Milton et al. 1995, Sandmeier 2000), as outlined below:

- It can be used to characterise the reflectance of GTs used for atmospheric correction (e.g., for the ELM) or vicarious calibration (Slater et al. 1987, Gu et al. 1992, Milton et al. 1995) and to validate the outcome of RT models.
- It provides a tool to examine relationships between biophysical variables and reflectance factors (Milton 1987, Milton et al. 1995, Sandmeier 2000).
- It acts as a bridge between laboratory measurements of reflectance and the field situation (Milton 1987).
- It can be used to investigate the optimum spectral bands and viewing conditions for a remote sensing task (Milton 1987). Hence, it can provide a relatively low cost pilot study.
- It can be used to validate airborne or satellite-inferred BRDF data (Sandmeier 2000).
- It can be used to support the development of BRDF models (Milton et al. 1995, Sandmeier & Itten 1999, Sandmeier 2000).
- It can be used to calibrate large and small reflectance reference panels (Sandmeier 2000, Rollin et al. 2000).
- It can be used as a remote sensing tool in its own right (Milton et al. 1995, Forster & Jesus 2006).

For the purpose of this thesis the use of field spectroscopy for characterising GTs was of central importance.

Overviews of field spectroscopy are provided by Milton and colleagues (Milton 1987, Milton et al. 1995) with more recent reviews focusing on BRDF measurements (e.g., Sandmeier 2000) and the use of low cost spectrometers (e.g., Forster & Jesus 2006). A detailed review is not provided here, although key issues are raised with respect to the use of the Milton Multiband Radiometer (MMR), which was used for the fieldwork conducted for this thesis.

4.3.1 The Milton Multiband Radiometer

The Milton Multiband Radiometer was initially developed at Reading University, UK by Professor E.J. Milton in the late 1970s and has continued to be developed since then. The instrument was designed as a low cost radiometer for teaching and research. Descriptions and evaluations of early versions have been published (Milton 1980, Milton 1981b) and subsequent revisions are documented in the grey literature (Milton 2001)[7]. Milton (2001) constitutes a user manual for the instrument.

The radiometer is comprised of two key units: the sensor unit and the display unit or data logger. The sensor unit contains four separate silicon photodiode detectors and each detector has one or more gelatin filters, as required to produce the appropriate filter–sensor spectral response. The sensor unit also contains four modular pre–amplifiers. The spectral response of the original MMR was selected on the basis of the first four waveband of the Landsat Multispectral Scanner (MSS). The hood in front of the MMR is used to set the sensor field of view (FOV). Note also that the gain of the instrument can be adapted.

The sensor unit is connected to a custom–built data logger or digital display. The instrument can be used in single beam or dual beam mode. If the instrument is operated in dual beam mode then the simultaneous measurements of the target and reflectance panel can be made. The unit gives an output in digital number (DN) proportional to the radiant intensity. The DNs are not calibrated to radiance, since the instrument was designed to be used with a calibrated reference panel.

The spectral bandpass of the instrument used is shown in table 4.7, together with the corresponding ATM bands. The spectral response of the two sensor heads used is shown in figure 4.3. It is clear from this figure that the spectral response of the two sensor heads are closely matched and are appropriate for use in dual beam mode (Milton 1980, Milton 1981a, Milton 1982, Duggin & Philpson 1982, Milton 1987).

Figure 4.4 shows the spectral response of the MMR and the most closely related ATM bands. It is clear that the MMR and ATM bands are not matched precisely. For implementation of the ELM or for validation of atmospheric correction it would be ideal to use either a precisely matched radiometer or use a spectrometer and convolve the output spectra with the spectral response of the airborne instrument. Neither option was available for this project. However, the MMR is useful because it is straightforward and fast to operate and was shown to give consistent results. Hence, it was appropriate for collecting a spatially referenced large sample and for tackling the key questions raised in this thesis. Band 2 of the ATM is similar to band 1 of the MMR, although the peak is slightly offset. The same conclusion is drawn for band 3 of the ATM and band 2 of the MMR. The spectral response of band 3 of the MMR lies between bands 4 and 5 of the ATM. It is more similar to band 4, but slightly offset from it. However, over the GTs, bands 4 and 5 are highly correlated (> 0.99). Hence band 3 of the MMR was linked to

[7]See also Professor Milton's webpage: `http://www.soton.ac.uk/~ejm/MMR/`

MMR band	Wavelength μm	ATM band	Wavelength μm
1	0.43 – 0.54	2	0.45 – 0.52
2	0.51 – 0.61	3	0.52 – 0.60
3	0.60 – 0.70	4 & 5	0.60 – 0.62 (4) & 0.63 – 0.69 (5)
4	0.72 – 1.13	7 & 8	0.76 – 0.90 (7) & 0.91 – 1.05 (8)

TABLE 4.7: Table showing the MMR wavebands for instruments SOT72 and SOT73 (data provided by the EPFS technician). The closest ATM wavebands are also shown (Source: NERC ARSF).

band 5 of the ATM for the empirical line method. Band 4 of the MMR is much wider than the ATM bands. It spans band 8 of the ATM, but also encompasses some of band 7, although ATM bands 7 and 8 are also highly correlated. In the analysis chapters band 4 of the ATM was linked to band 7 of the ATM.

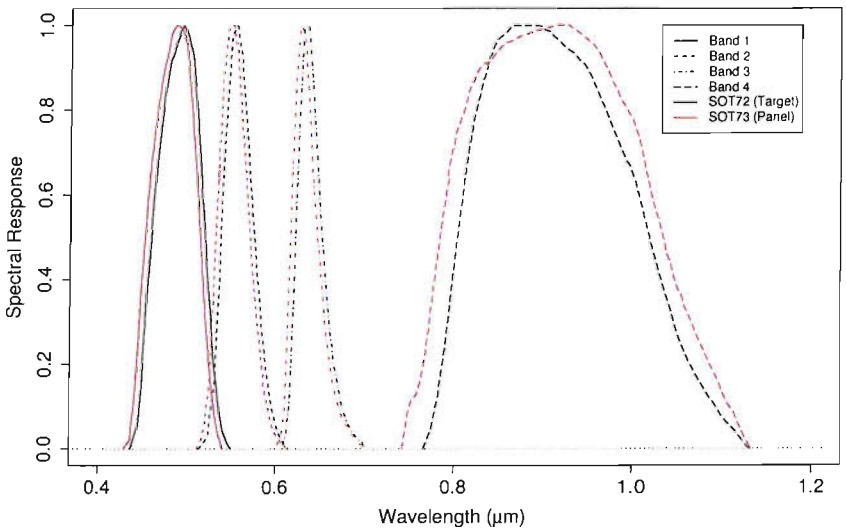


FIGURE 4.3: Figure showing the spectral response of the two MMR heads used for this project (data provided by the EPFS technician). SOT72 is designated for viewing the target and SOT73 is designated for viewing the panel.

For this work, the instrument was operated in dual beam mode. The reflectance factor is given as:

$$\rho_\lambda = \frac{DN'_{T_\lambda}}{DN'_{P_\lambda}} k_\lambda C_\lambda \tag{4.1}$$

$$DN'_{T_\lambda} = DN_{T_\lambda} - DC_{T_\lambda} \tag{4.2}$$

$$DN'_{P_\lambda} = DN_{P_\lambda} - DC_{P_\lambda} \tag{4.3}$$

where:

ρ is the reflectance factor, given a as a fraction between 0 and 1 (alternatively expressed as a percentage);

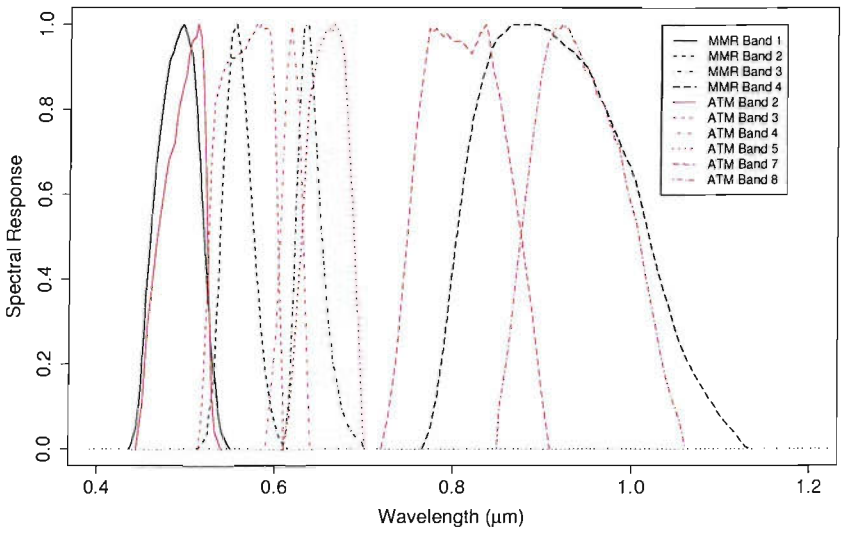


FIGURE 4.4: Figure showing the spectral response of the MMR (SOT72) and the ATM bands that match most closely.

DN is the digital number;

DC is DN of the dark level offset;

T refers to the MMR designated to measure the radiance reflected from the target;

P refers to the MMR designated to measure the radiance reflected from the panel;

k is the panel correction factor;

C is the inter-calibration coefficient;

λ refers to the waveband.

To derive ρ it is necessary to establish k and C. The latter is established by inter-calibrating the two radiometers (Milton 1981a, Duggin & Philpson 1982, Milton 1987, Milton et al. 1995). The two radiometers used for this project are of the same specification, have the same filters fitted and have similar spectral responses. They are, therefore, well suited for use in dual beam mode.

To collect the data required for the inter-calibration the panel was set up on a tripod and the two sensor heads were attached, so that both viewed the panel. The heads were attached to the data logger and set to record simultaneous dual beam measurements at short time intervals (up to 1 minute). The instruments were positioned in a wide open space that was free from shadow. Where possible, the instruments were placed away from tall buildings or trees, to minimise the obstruction of skylight. The inter-calibration exercise was repeated on several occasions either at Southampton University campus, on Southampton Common, at Thorney Island or in the New Forest. The inter-calibration

data sets were collected under a range of illumination conditions, including bright days with intermittent clouds, as suggested by Milton (2001). Collecting inter-calibration data sets on several dates is useful, since it also provides an indication of whether there is any drift in the relative spectral response of either instrument.

If the data collected from each instrument are plotted against each other, then the coefficient, C_λ, can be determined by the gradient of the line given by (Duggin & Philpson 1982, Milton 1987, Milton 2001):

$$C_\lambda = \frac{DN'_{P_\lambda}}{DN'_{T_\lambda}} \tag{4.4}$$

However, equation 4.4 requires that all the data points lie on a straight line that passes through the origin. In reality, this is not the case and Duggin & Philpson (1982) and Milton (2001) use regression to calculate C. Hence C is obtained from the slope of the regression line (i.e., $C = \hat{\beta}_1$).

Milton (2001) recommends that, for any given band, the correlation between the data set measured for each instrument should be very high ($R^2 \geq 0.99$) , to ensure a high quality inter-calibration. He then recommends that C is found by regressing DN'_P on DN'_T and taking C as the slope of the regression line (i.e., $C = \hat{\beta}_1$).

Using regression to calculate C is useful, since it recognises that there the data points may deviate from a straight line, and provides a mechanism for plotting a best fit line through the data. The data may deviate from a straight line for various reasons, including noise in either of the sensors or short–term spatial variation in illumination of the panel. However, as discussed in chapter 2 simple linear regression is predicated on the case that the independent variable, DN'_T, is known without error. This is *not* the case in this situation since both DN'_T and DN'_P are subject to error. This is illustrated as follows:

$$DN'_{P_i} = DN'^{*}_{P_i} + \varepsilon_i \tag{4.5}$$

$$DN'_{T_i} = DN'^{*}_{T_i} + \eta_i \tag{4.6}$$

where the $*$ indicates the true value of DN' and i refers to a particular measurement. Dependence on waveband is suppressed. Hence, the regression equation:

$$DN'_{P_i} = \beta_0 + \beta_1 DN'^{*}_{T_i} + \varepsilon_i \tag{4.7}$$

becomes:

$$DN'_{P_i} = \beta_0 + \beta_1 (DN'_{T_i} + \eta_i) + \varepsilon_i \tag{4.8}$$

which violates the basis of simple linear regression (Sprent 1969, Curran & Hay 1986, Webster 1989, Neter et al. 1996). Further discussion is provided in chapter 2 and by Kendall & Stuart (1967).

This issue can be addressed if it is assumed that ϵ and η have the following simple

properties (Sprent 1969, Kendall & Stuart 1967):

$$\begin{aligned} \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \text{ and } \eta_i \sim N(0, \sigma_\eta^2), \text{ for all } i; \\ Cov(\varepsilon_i, \varepsilon_j) = Cov(\eta_i, \eta_j) = 0,\ i \neq j; \\ Cov(\varepsilon_i, \eta_j) = 0, \text{ for all } i,\ j. \end{aligned} \tag{4.9}$$

From here the maximum likelihood estimator of β_1 (i.e., $\hat{\beta}_1$) is given as:

$$\hat{\beta}_1 = \frac{\sum y_i^2 - \lambda \sum x_i^2 + \sqrt{\{(\sum y_i^2 - \lambda \sum x_i^2)^2 + 4\lambda(\sum x_i y_i)^2\}}}{2 \sum x_i y_i} \tag{4.10}$$

where $\lambda = \sigma_\varepsilon^2 / \sigma_\eta^2$. This formulation is achieved by setting the origin to $(\overline{DN_T}, \overline{DN_P})$ (hence $\beta_0 = 0$), so $x_i = DN_{T_i} - \overline{DN_T}$ and $y_i = DN_{P_i} - \overline{DN_P}$. For this analysis it was assumed that $\lambda = 1$. This is equivalent to minmising the variance perpendicular to the regression line which may be termed the *principal axis* (Webster 1989) and is termed PAR (principal axis regression) for this discussion. Setting $\lambda = 1$ is well founded since the instruments are designed, built and tested to the same specification and the measurements from both instruments are made under the same conditions.

An inter-calibration data set was collected on 20th July 2003 and information about this is given in table 4.8. The results shown in table 4.9 show the point estimates of β_0, β_1 and σ^2 using OLS and PAR, as well as the value for the correlation coefficient, r. The result of t–tests show that, for both OLS and PAR, $\hat{\beta}_0 \neq 0$ and $\hat{\beta}_1 \neq 0$ at the 95% confidence level. The estimates of β_0 and β_1, gained from the two regression methods are very similar. The estimate of σ^2 is much lower for PAR than for OLS, which also means that confidence intervals around the parameter estimates and new predictions ($\hat{y}$) are lower. If a practitioner was to use OLS rather than PAR the practical consequence would, in this case, be limited, since the estimates of β_0 and β_1 are so similar. This is demonstrated if predictions based on $\hat{\beta}_0$ and $\hat{\beta}_1$ gained using the two techniques (denoted $\hat{y}_{OLS}$ and $\hat{y}_{PAR}$) are compared. For band 1, $\hat{y}_{OLS} - \hat{y}_{PAR} < 0.003$, for band 2 $\hat{y}_{OLS} - \hat{y}_{PAR} < 0.007$, for band 3 $\hat{y}_{OLS} - \hat{y}_{PAR} < 0.003$, for band 4 $\hat{y}_{OLS} - \hat{y}_{PAR} < 0.008$. This is several orders of magnitude lower than the value of y, which is in the region 10–1000.

Having implemented the regression model and estimated the parameters it is clear that $\hat{\beta}_1$ from equation 4.7 or 4.8 corresponds to C_λ from equations 4.1 and 4.4. However, the method given in equation 4.1 does not incorporate β_0. This implicitly assumes that $\beta_0 = 0$, which is what would be expected for an instrument that has a stated linear response to the intensity of radiation. However, table 4.8 shows that $\hat{\beta}_0 \neq 0$, which is a source of concern. This implies that there is some bias in the instrument response that is not corrected for by subtracting the dark current. There may also be some non–linearity in the instrument response at low irradiance values (i.e., below the minimum value collected for the inter-calibration).

In view of the fact that $\hat{\beta}_0 \neq 0$ it might be argued that equation 4.1 should be re–written

Date	20th July 2001
Time	13:25 – 15:15 (GMT)
Location	Southampton Common
Weather	Clear blue skies with intermittent cloud
Instruments	SOT-72H (Target) and SOT-73H (Surface)

Band		No. of measurements	Min	Max
Band 1	$DN_T\ (x)$	631	42	538
	$DN_P\ (y)$	631	46	473
Band 2	$DN_T\ (x)$	410	102	846
	$DN_P\ (y)$	410	35	302
Band 3	$DN_T\ (x)$	631	33	491
	$DN_P\ (y)$	631	31	466
Band 4	$DN_T\ (x)$	631	26	431
	$DN_P\ (y)$	631	32	519

TABLE 4.8: Information about the inter-calibration data set collected on 20th July 2001.

		OLS			PAR		
MMR Band	r	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}^2_\varepsilon = \hat{\sigma}^2_\eta$
Band 1	$r > 0.99$	-1.2035	0.8815	0.7268	-1.2055	0.8816	0.4090
Band 2	$r > 0.99$	-1.4185	0.3592	0.3660	-1.4204	0.3592	0.3242
Band 3	$r > 0.99$	-1.0840	0.9517	0.7578	-1.0861	0.9517	0.3976
Band 4	$r > 0.99$	0.7644	1.2112	1.5573	0.7596	1.2112	0.6312

TABLE 4.9: The estimates of β_0, β_1 and σ^2 using ordinary least squares (OLS) and principal axis regression (PAR). The value of the correlation coefficient, r, is also shown. This data set was collected on 20th July 2001.

as:

$$\rho_\lambda = \frac{\hat{\beta}_0 + \hat{\beta}_1 DN'_{T_\lambda}}{DN'_{P_\lambda}} k_\lambda \tag{4.11}$$

The difference between adopting equations 4.1 and 4.11, was assessed by applying the two models to dual–beam measurements made over asphalt, concrete and grass. The results are shown in table 4.10. These are the data sets used for the analysis in chapters 5 and 7. Equation 4.11 is problematic, since routine data collection showed that $DN'_{T_\lambda} = 0$ when $DN'_{P_\lambda} = 0$. In view of this, and the fact that equation 4.1 is widely referred to in the literature (e.g., Duggin & Philpson 1982, Milton 1987, Peddle et al. 2001) and in the instruction manual (Milton 2001) it was decided to adopt equation 4.1. The issues raised above are noted with concern, but the following caveats are made

1. The MMR is a low-cost instrument. Although it is reliable and has been widely used, the limitations of the instrument need to be accepted.

2. The inter-calibration data set was gathered over a range of solar irradiance values. The solar irradiance varies, both in intensity and distribution. However, when the dual beam measurements are implemented the solar irradiance remains fixed

Band 1	$0.2 <$	$\frac{\hat{\beta}_1 DN_T}{DN_P} - \frac{\hat{\beta}_0 + \hat{\beta}_1 DN_T}{DN_P}$	< 0.6
Band 2	$0.2 <$	$\frac{\hat{\beta}_1 DN_T}{DN_P} - \frac{\hat{\beta}_0 + \hat{\beta}_1 DN_T}{DN_P}$	< 0.8
Band 3	$0.2 <$	$\frac{\hat{\beta}_1 DN_T}{DN_P} - \frac{\hat{\beta}_0 + \hat{\beta}_1 DN_T}{DN_P}$	< 0.7
Band 4	$-0.5 <$	$\frac{\hat{\beta}_1 DN_T}{DN_P} - \frac{\hat{\beta}_0 + \hat{\beta}_1 DN_T}{DN_P}$	< -0.1

TABLE 4.10: Difference between adopting the inter-calibration model given in equation 4.1 or 4.11, for the data set collected on 20th July 2001. The units are reflectance (%). These upper and lower bounds apply to all three surfaces (asphalt, concrete and grass).

(between DT_P and DT_T) and the reflectance of the surface varies. Notwithstanding the bidirectional properties of the target, the inter-calibration procedure implicitly assumes that these induce the same effect, where in reality, they may not. Indeed, this is one argument for using a reflectance panel that has a similar reflectance value to the target of interest (Duggin & Philpson 1982, Philipson et al. 1989).

The above discussion has demonstrated how the inter-calibration was carried out for one date. It should be noted that, if there is any drift in the instrument over time this may lead to a change in the inter-calibration parameters. Ideally, one would want to inter-calibrate the instrument each time it is used. However, an inter-calibration data set needs to be collected under a range of illumination conditions. Therefore, conditions on the day of use may not be suitable and, in any case, the priority will be to collect field data. In practice, it was not straightforward to find days where the weather was suitable and the facilities were available to make inter-calibration measurements. Hence, if field inter-calibration data sets could not be collected on the day of use every effort was made to ensure that measurements were taken the day before use and the day after use.

For this project, MMR measurements were made in June, July, September, October and November. The key data set was collected on the 23rd, 24th and 25th July. Several inter-calibration data sets were taken on various days from May to November. Examples are shown for six dates on 20th June, 20th July, 26th July, 5th October and 6th November 2001 in table 4.11. This table demonstrates the importance of checking the validity of the inter-calibration for each set of measurements. It is important to note that the inter-calibration data sets collected on 20th and 26th July give consistent estimates of the β_1 parameter. The key data set was collected on the 23rd, 24th and 25th of July. To test the impact of the variability in the inter-calibration parameter estimates (equation 4.1) reflectance from the asphalt, concrete and grass data sets was calculated using estimates given in table 4.11. The reflectance values predicted using $\hat{\beta}_1$ values calculated on 20th July were used as a benchmark and the reflectance predictions using $\hat{\beta}_1$ values from other dates were compared to these. The maximum differences were 0.1% (reflectance units) for band 1; 0.12% for band 2; 1% for band 3 and 1% for band 4. Using the inter-calibration data sets collected on 20th June, 20th July and 26th July would lead to a difference in predicted reflectance of no more than 0.1% in any band for any surface.

MMR Band	Parameter Estimate	20/06/01	20/07/01	26/07/01	05/10/01	06/11/01
Band 1	$\hat{\beta}_0$	-1.1863	-1.2055	-1.0897	-0.6504	-0.6711
	$\hat{\beta}_1$	0.8815	0.8816	0.8811	0.8784	0.8773
Band 2	$\hat{\beta}_0$	-1.2880	-1.4204	-1.2230	-1.8199	-1.1864
	$\hat{\beta}_1$	0.3585	0.3592	0.3585	0.3576	0.3579
Band 3	$\hat{\beta}_0$	-0.9285	-1.0861	-0.9285	-1.4838	-0.4294
	$\hat{\beta}_1$	0.9512	0.9517	0.9518	0.9380	0.9860
Band 4	$\hat{\beta}_0$	0.8533	0.7596	0.8665	2.5790	0.3071
	$\hat{\beta}_1$	1.2108	1.2112	1.2113	1.1853	1.1973

TABLE 4.11: Parameter estimates (PAR) for inter-calibration data sets collected on 20th June, 20th July, 26th July, 5th October and 6th November 2001.

Finally, it is reasonable to question why equation 4.1 should be adopted in preference to equation 4.12:

$$\rho_\lambda = \frac{DN'_{T_\lambda}}{\beta_1 DN'_{P_\lambda}} k_\lambda \tag{4.12}$$

or that equation 4.13 should be considered rather than equation 4.11

$$\rho_\lambda = \frac{DN'_{T_\lambda}}{\beta_0 + \beta_1 DN'_{P_\lambda}} k_\lambda \tag{4.13}$$

In both of these cases the regression parameters are dealt with in the denominator rather than following the convention of dealing with them in the numerator. There are good reasons for considering use of equation 4.12, as follows:

1. When equation 4.1 is applied to dual beam measurements, the values of DN'_{T_λ} often lie outside the range of values that were used to estimate $\hat{\beta}_1$ in the inter-calibration process. This is very possible if the surface is dark. This amounts to extrapolation. This issue tends not to arise if equation 4.12 is adopted. However, although this issue is important from an theoretical perspective, it makes no practical difference to the prediction of ρ_λ.

2. The difference between using equation 4.11 rather than equation 4.1 was discussed previously. However, if the choice between equations 4.12 and 4.13 is considered, the decision is relatively unimportant, since the inclusion of β_0 makes a difference of less than 0.01 % in the prediction of ρ_λ. This is simply due to the mathematics, and the effect can be demonstrated using a simple example: $a = 10/100 = 0.1$, $b = (10 + 1)/100 = 0.11$ and $c = 10/(100 + 1) = 0.0990099$, $|a - c| \ll |a - b|$.

3. Where β_1 varies over time, the effect of variability in β_1 on the prediction of ρ_λ is the same whether equation 4.1 or equation 4.12 is used. This is can be established using algebra.

The conclusion here is that, if the decision is made to omit the β_0 parameter (i.e., use

Panel Serial Number	srt7165	
MMR Head (P)	SOT 973H	
MMR Head (T)	SOT 972H	
Band	k (SOT 973H)	k (SOT 972H)
Band 1	0.9995	0.9895
Band 2	0.9900	0.9900
Band 3	0.9895	0.9895
Band 4	0.9900	0.9900

TABLE 4.12: Panel correction factors for the MMR wavebands.

equation 4.1 or 4.12 rather than equation 4.11 or 4.13), then equation 4.1 and equation 4.12 can be used interchangeably. However, if the β_0 parameter is included, then equation 4.13 is preferable to equation 4.11.

Having inter–calibrated the two MMR heads, it is necessary to establish k in equation 4.1. This is achieved by convolving the spectral response of the MMR with the spectral response of the panel. One panel, loaned from the NERC EPFS, was used throughout this project. The EPFS provide data on the spectral response of each panel. This is convolved with the spectral response of the sensor to determine k for each waveband. The values of k are shown in table 4.12. Having ascertained this, all the information was available for implementing equation 4.1.

4.4 Field sampling

The aims and objectives for the thesis together with the key issues for investigation were introduced in section 1.3. Detailed background to ELM was provided in chapter 3.

To address those issues field sampling was focused on obtaining a dense spatial sample of the reflectance of the GTs. The GTs selected at Thorney Island were asphalt, concrete and cropped grass as adopted by Smith & Milton (1999). Targets were selected that were as free as possible from obvious discoloration or morphological features, such as grooves and bumps (Staenz & Itten 1982) and the ones chosen are illustrated in figure 4.2. It may be noted that the concrete sections of the site are laid in blocks of approximately 6 m $\times$ 6 m. Previous research had demonstrated that spatial structure existed in the nadir reflectance (visible and NIR) of all three cover types (Salvatori et al. 2000) and advocated gathering a spatially structured sample to characterise such targets.

Salvatori et al. (2000) utilised a transect, with measurements being taken at 1.5 m spacing. In reality, a 2-dimensional grid is needed to sample a GT. Hence nested grids were set up to cover a 32 m $\times$ 32 m area, with nadir reflectance measurements being taken at 1, 2 and 4 m intervals with the location of every measurement being recorded. This strategy was adopted as a trade-off between obtaining a *dense* and *spatially extensive* sample within an approximately two hour time window. This was found to be feasible during pilot studies

undertaken in early summer 2001. The two hour time window can be chosen to span solar noon, hence reducing the impact of BRDF effects.

As mentioned in section 4.2, the key ARSF flight undertaken for this thesis was undertaken on 24 July 2001. Hence, the GTs were sampled around solar noon on 23rd July 2001 (asphalt), 24 July 2001 (grass) and 25 July 2001 (concrete). Ideally, the targets should be sampled as close to the flight as possible (Smith & Milton 1999). However, an unfortunate consequence of undertaking a spatial sample is that it is time-consuming. Furthermore, it was not possible to sample the targets during the flight window. Hence the procedure adopted was the only one feasible. Furthermore, it may be noted that the weather remained stable over these time-periods and there were no obvious visual changes in the surfaces.

Recent research (Anderson & Milton 2005, Anderson & Milton 2006) has demonstrated that the reflectance of targets can change over short time-scales. However, this is not an issue that can be addressed in this thesis. Linking spatial and temporal sampling and modelling remains an avenue for future research.

The two MMR instruments were set up on a mast with one instrument (SOT72) pointing towards the ground surface with the second instrument (SOT73) oriented over a spectralon panel. When the mast is held vertically, both instruments are at nadir and SOT72 is 1.9 m above the ground. A spirit level on the mast assists the user in accurately positioning the mast vertically. Since the instrument has a FOV of approximately 15°, the MMR images an area on the ground of 0.50 m in diameter. However, given the point-spread function (PSF) (unpublished data provided by the EPFS technician) of the MMR, it is expected that most of the signal should result from the a central area of approximately 0.25–0.30 m in diameter. The instruments were connected to a data-logger which controlled both instruments simultaneously and recorded the DN from both instruments in each waveband. These DNs were then processed to reflectance using the method established in section 4.3.1. The grids, together with summary statistics are presented in section 4.5.

It is *essential* to note that the location of every field measurement was marked and later surveyed using a total station. The location of other key points such as corners, land-marks and OS benchmarks were also surveyed. It is these points that were used in section 4.2.3 to assess the accuracy of the geometric correction. Because of the care taken to mark and survey the field measurements it was possible to determine accurately the location of *every* field measurement. Hence the location of the field measurements within the image was determined primarily by the geometric correction.

Having sampled the reflectance of the GTs, the next stage in implementation of the ELM was to combine these measurements with the remotely sensed radiance measurements. These issues were discussed in sections 3.2.3.4 and 3.2.3.5 and are addressed empirically in chapters 5 and 6.

Finally, note that the MMR wavebands better approximate the ATM than the CASI.

Hence, most of the analysis undertaken for this thesis was undertaken using the ATM data. However, chapter 6 draws heavily on CASI data. This is because this research was undertaken before the ATM data became available.

4.5 Data summaries

This section provides a general overview of the field and image data. The objective is to provide a clearer picture of the sampling approach and to provide a reference for subsequent analysis.

4.5.1 Field data

Summary statistics for the reflectance data gathered over the three surfaces are shown in tables 4.13 (asphalt), 4.14 (concrete) and 4.15 (grass). Summary plots for the reflectance data gathered using the MMR are shown in figures 4.5 (asphalt), 4.6 (concrete) and 4.7 (grass). For each surface (asphalt, concrete, and grass), the summary plots are grouped by waveband (blue, green, red, NIR). For each waveband the histograms and QQ plots (quantile–quantile plots) provide information about the aspatial characteristics of the data. The QQ plot is a simple visual tool that allows an informal assessment of whether the data are drawn from a given distribution (Chambers et al. 1983). For a sample drawn from a normal distribution, the data should lie along the 1:1 diagonal line. The histograms and QQ plots do not represent formal tests. Even data that are simulated from a normal distribution will not yield perfect histograms or QQ plots (Chambers et al. 1983). For each waveband, spatial summary plots are also shown. These show the magnitude of the data values in plan view (where the size of the point is proportional to the magnitude of the measurement), as well as two dimensional plots showing the data values against eastings and northings. The omnidirectional variogram is also shown.

For the asphalt surface, figure 4.5 shows that the data appear to be normally distributed in all bands. This is supported by the skewness and kurtosis statistics provided in table 4.13 (Webster & Oliver 2001). Where the data points are above the theoretical line on the right hand side and below it on the left hand side, this is indicative of long tails. The opposite is indicative of short tails (Chambers et al. 1983). There is some slight indication of long tails for bands 1 (blue) and 3 (red) and some slight indication of short tails for band 4 (NIR).

For the concrete surface, figure 4.6 shows that the data are not normally distributed in any of the bands. This is supported by the statistics provided in table 4.14, where the values for skewness and kurtosis suggest that the data are skewed from a normal distribution. This is particularly clear for bands 1 to 3. Analysis of the spatial distribution of the data values suggests that the concrete target can be split into two distinct sub–targets.

Waveband	Minimum	Median	Mean	Maximum	St. Dev	Skewness	Kurtosis
Band 1	7.91	8.57	8.57	9.40	0.27	0.24	0.44
Band 2	8.77	9.67	9.66	10.57	0.32	0.02	0.01
Band 3	9.26	10.37	10.37	11.59	0.37	-0.05	0.43
Band 4	12.3	13.53	13.51	14.90	0.57	-0.01	-0.62

TABLE 4.13: Summary statistics for the MMR data gathered over the asphalt surface. Units are reflectance expressed as a percentage.

Waveband	Minimum	Median	Mean	Maximum	St. Dev	Skewness	Kurtosis
Band 1	11.65	14.19	14.72	19.78	1.89	0.97	0.06
Band 2	16.27	19.61	19.95	25.26	1.92	0.68	-0.09
Band 3	18.09	22.14	22.53	28.49	2.18	0.56	-0.18
Band 4	22.86	27.64	27.92	33.37	1.95	0.32	-0.33

TABLE 4.14: Summary statistics for the MMR data gathered over the concrete surface. Units are reflectance expressed as a percentage.

For the grass surface (figure 4.7) the data are approximately normally distributed. This is supported by the skewness and kurtosis statistics provided in table 4.15. There is some indication of asymmetry in bands 2, 3 and 4.

The spatial summary plots give a visual indication of the presence of spatial structure. For the asphalt surface (figure 4.5) the plots also show that points with similar values tend to be located close together. There is also a general indication of a decrease in the value of reflectance in all four bands moving from west to east. The variograms also provide evidence of spatial structure in all wavebands.

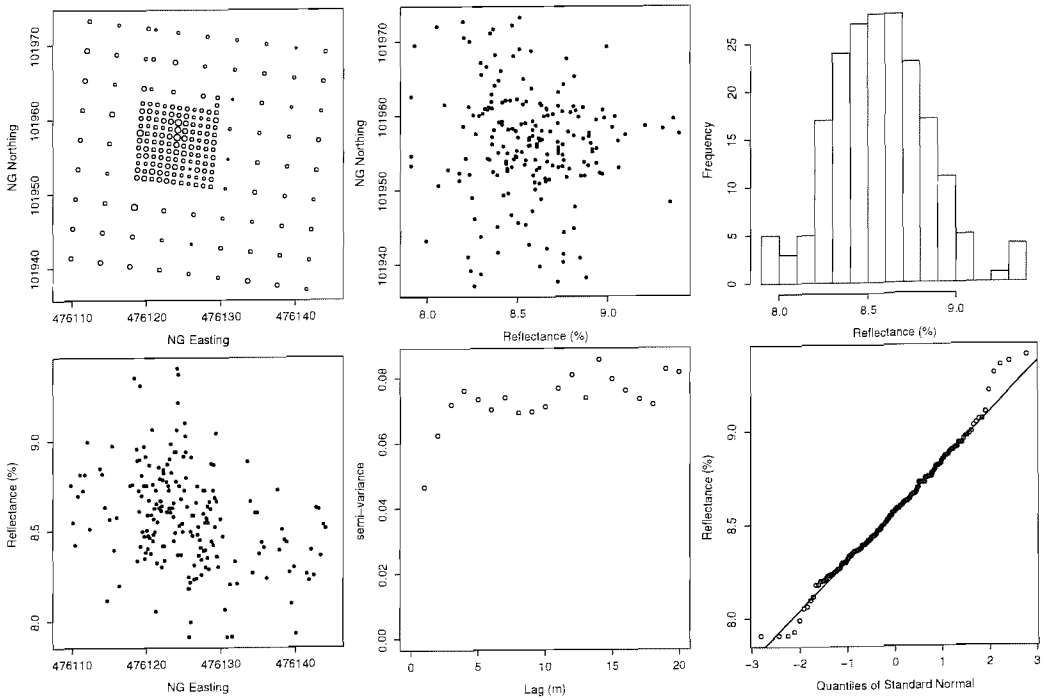
For the concrete surface (figure 4.6) the plots show that points with similar values tend to be located close together. There is no clear indication of a trend in the west–east or south–north direction. However, the north east quadrant appears to contain a cluster of higher values. This may account for the bimodal distribution in band 1 (figure 4.6(a)) and negatively skewed distributions for bands 2 to 4 (figures 4.6).

For the grass surface (figure 4.7) the plots show that points with similar values tend to be located close together. There is a clear indication of a trend, increasing from south to north. There also seems to be a strip of larger values on the north west edge in all four bands.

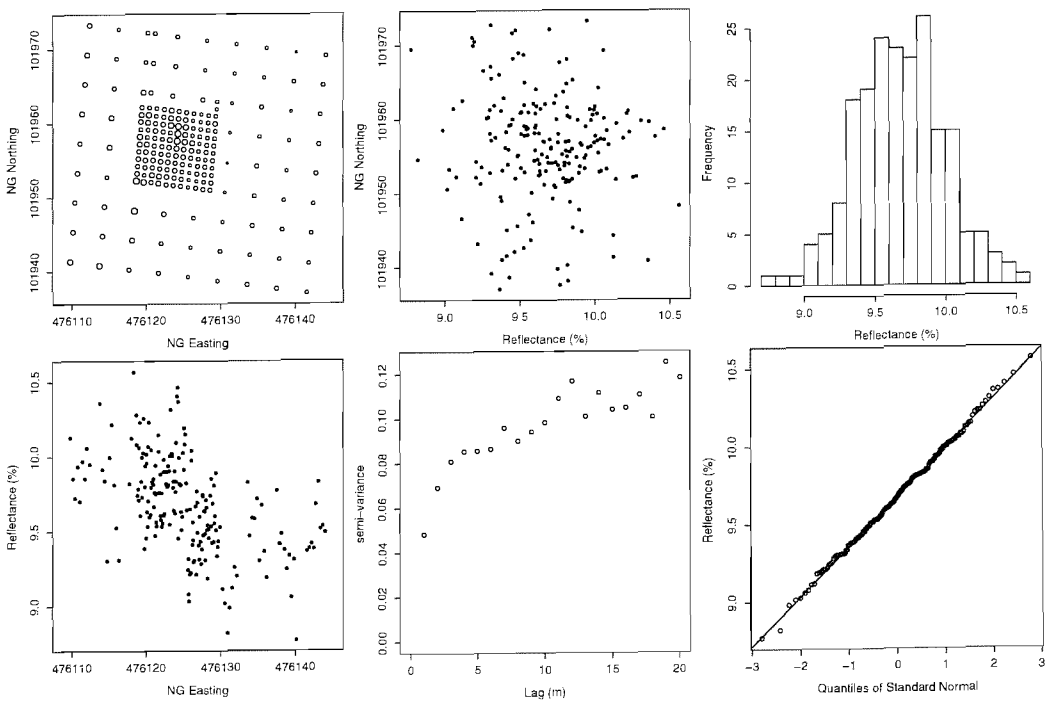
Figures 4.5, 4.6 and 4.7 show that ensembles of reflectance measurements that are made at the same location but in different wavebands can exhibit different distributions and different forms of spatial structure. A similar point is noted by Atkinson & Emery (1999) who observe different forms of spatial structure at different wavelengths for reflectance measurements made over heathland.

Waveband	Minimum	Median	Mean	Maximum	St. Dev	Skewness	Kurtosis
Band 1	2.53	3.69	3.70	5.00	0.44	0.19	-0.10
Band 2	5.39	7.46	7.47	9.18	0.34	-0.26	0.14
Band 3	4.76	6.14	6.65	9.08	0.81	0.32	-0.14
Band 4	26.09	34.4	34.40	39.74	2.26	-0.35	0.44

TABLE 4.15: Summary statistics for the MMR data gathered over the grass surface. Units are reflectance expressed as a percentage.

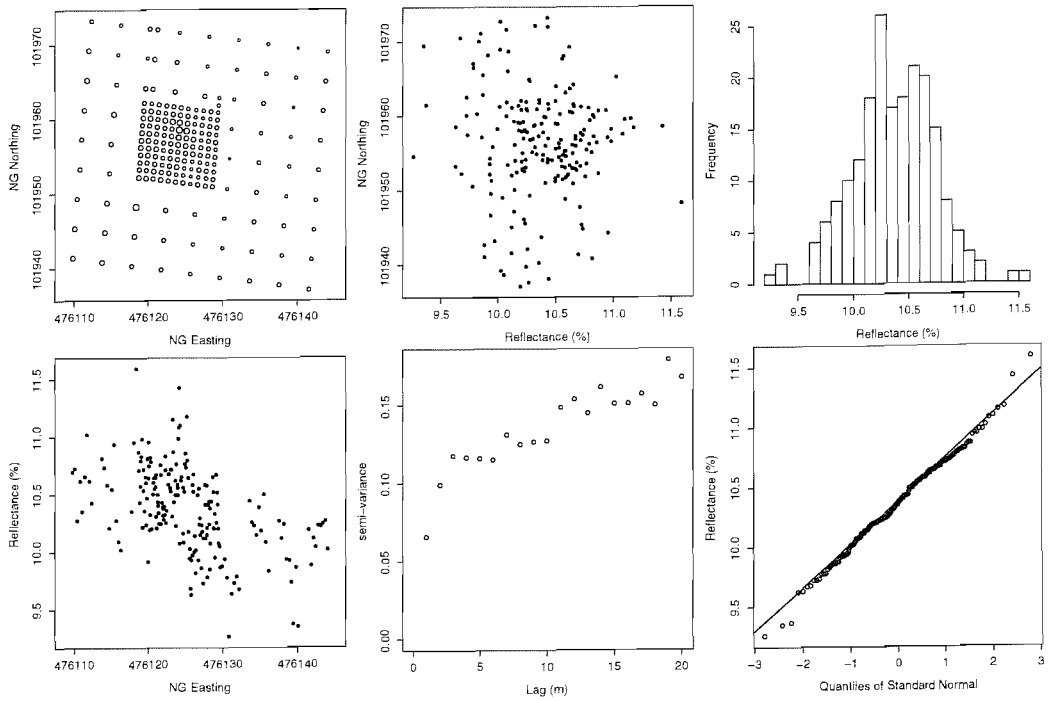


(a) MMR Band 1

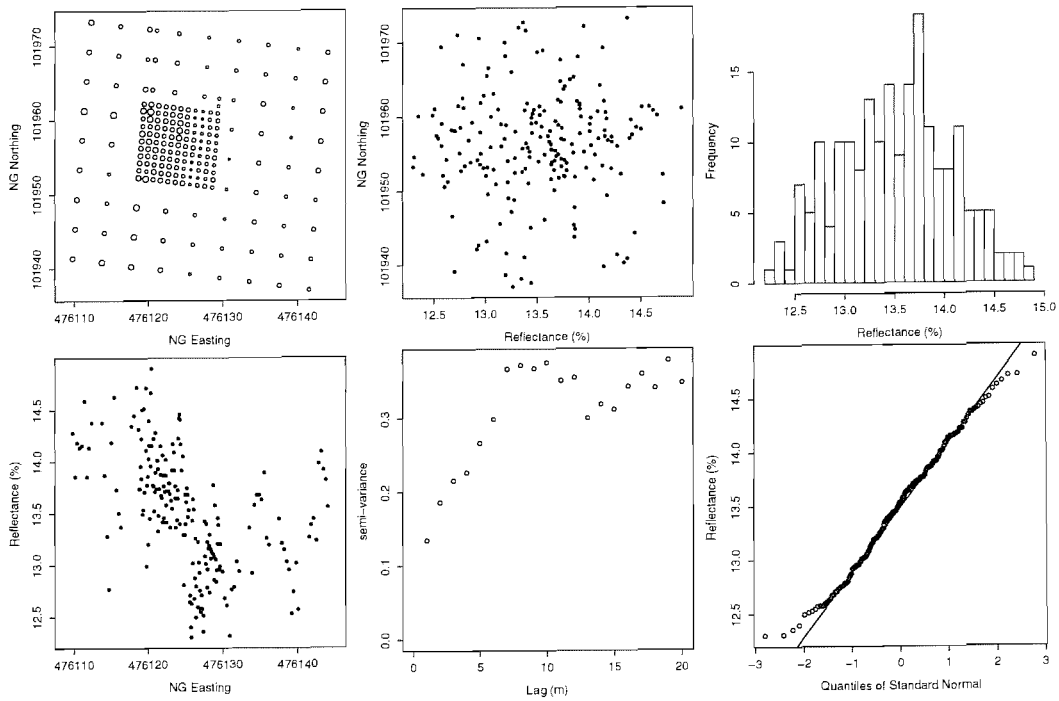


(b) MMR Band 2

FIGURE 4.5: Spatial summary plots for the MMR reflectance measurements taken over the asphalt surface. In the top left plot (in each sub-figure), the size of the circles is proportional to the magnitude of the reflectance value. The top middle (bottom left) plots show reflectance against easting (northing). The top right shows the histogram of data values and the bottom right the QQ plot. The bottom middle plot shows the omnidirectional sample variogram.

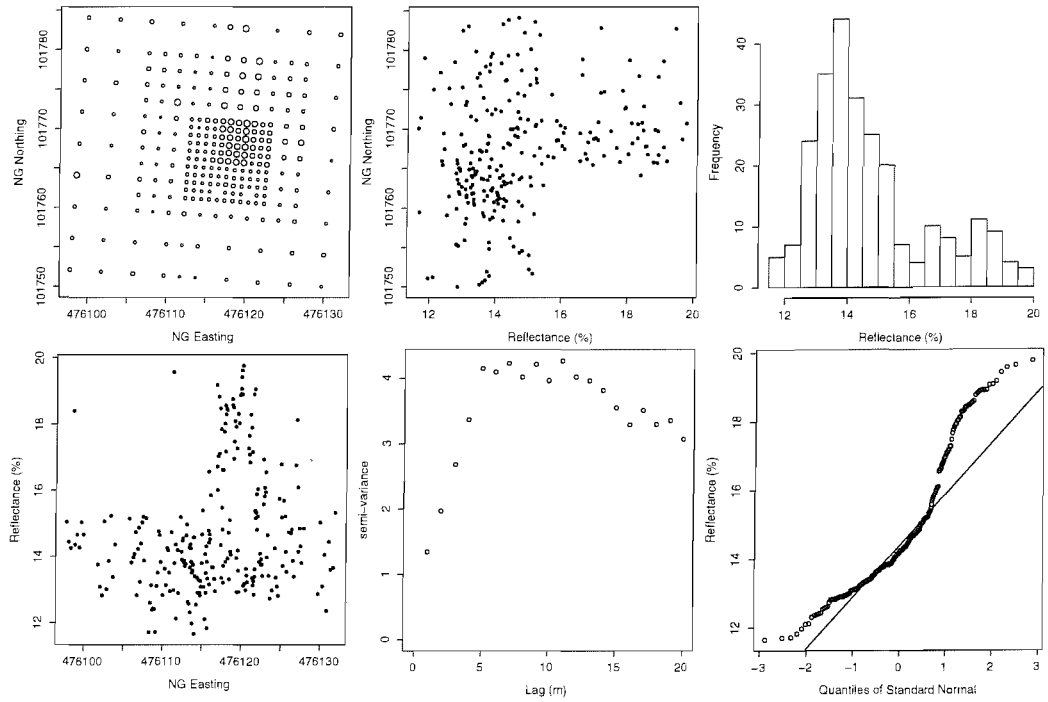


(c) MMR Band 3

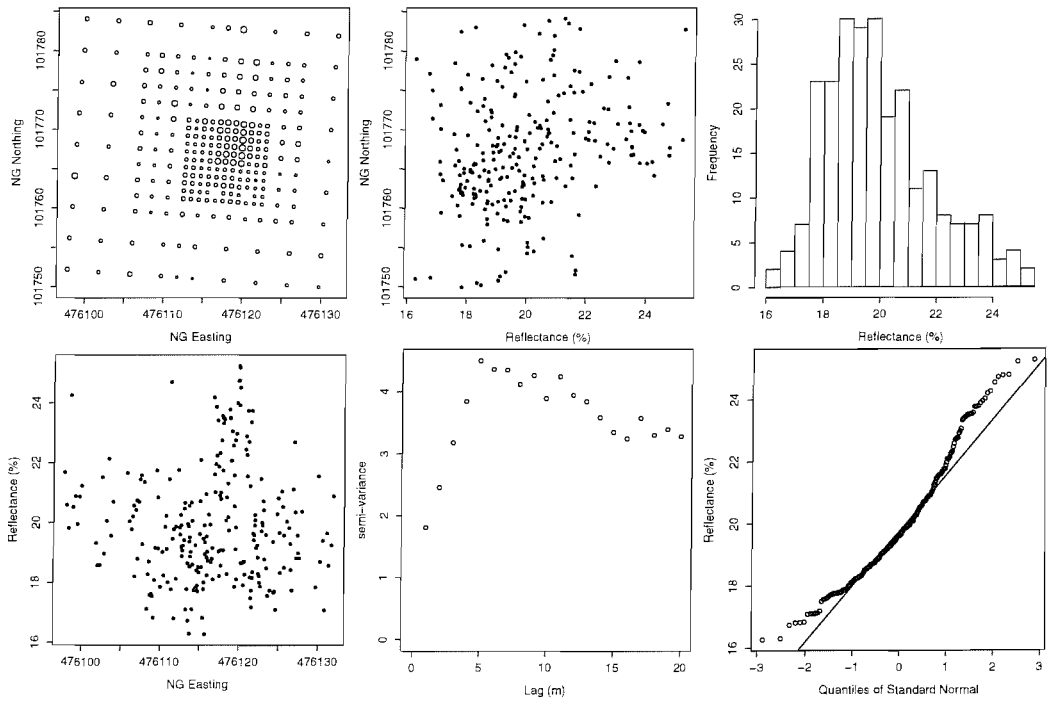


(d) MMR Band 4

FIGURE 4.5: continued

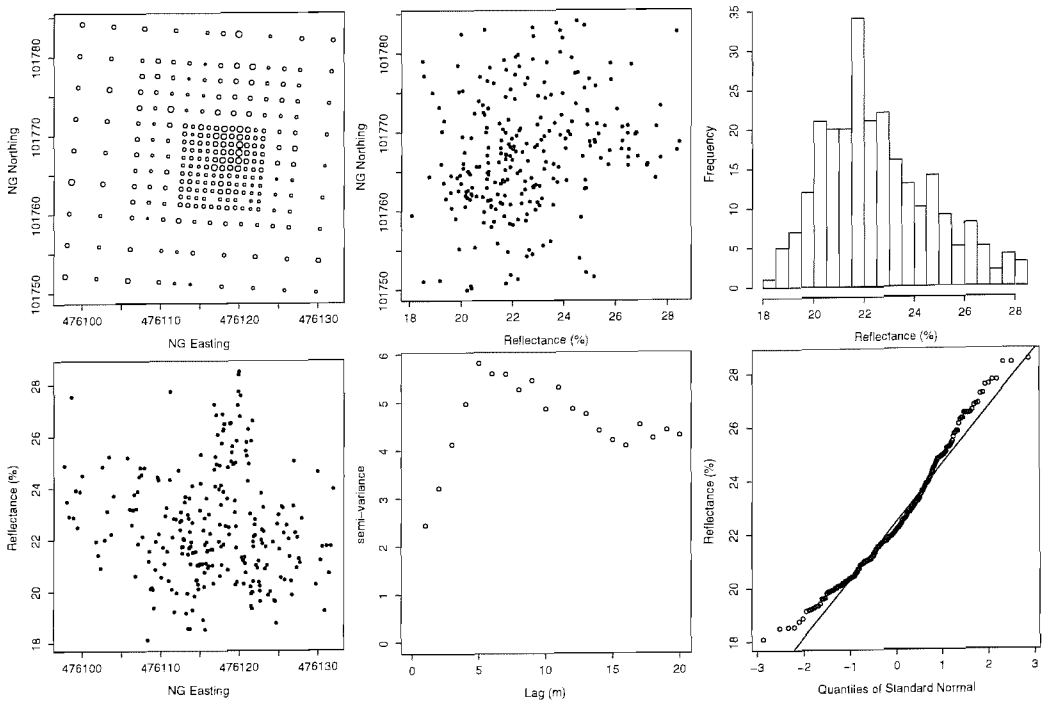


(a) MMR Band 1

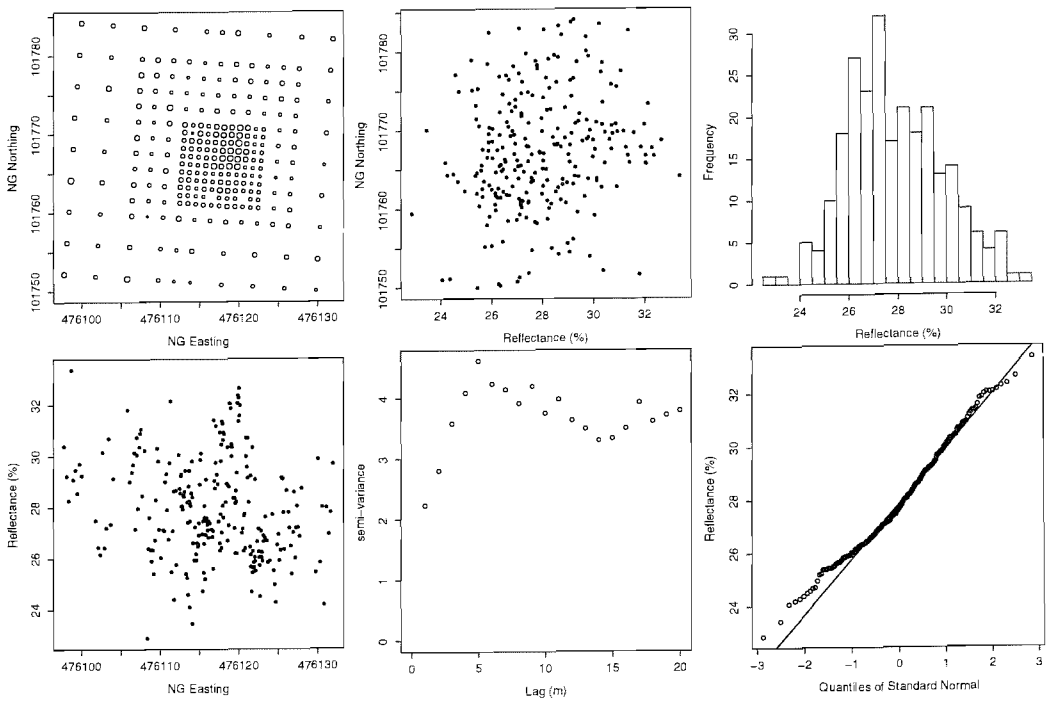


(b) MMR Band 2

FIGURE 4.6: Spatial summary plots for the MMR reflectance measurements taken over the concrete surface. Details are given in figure 4.5.

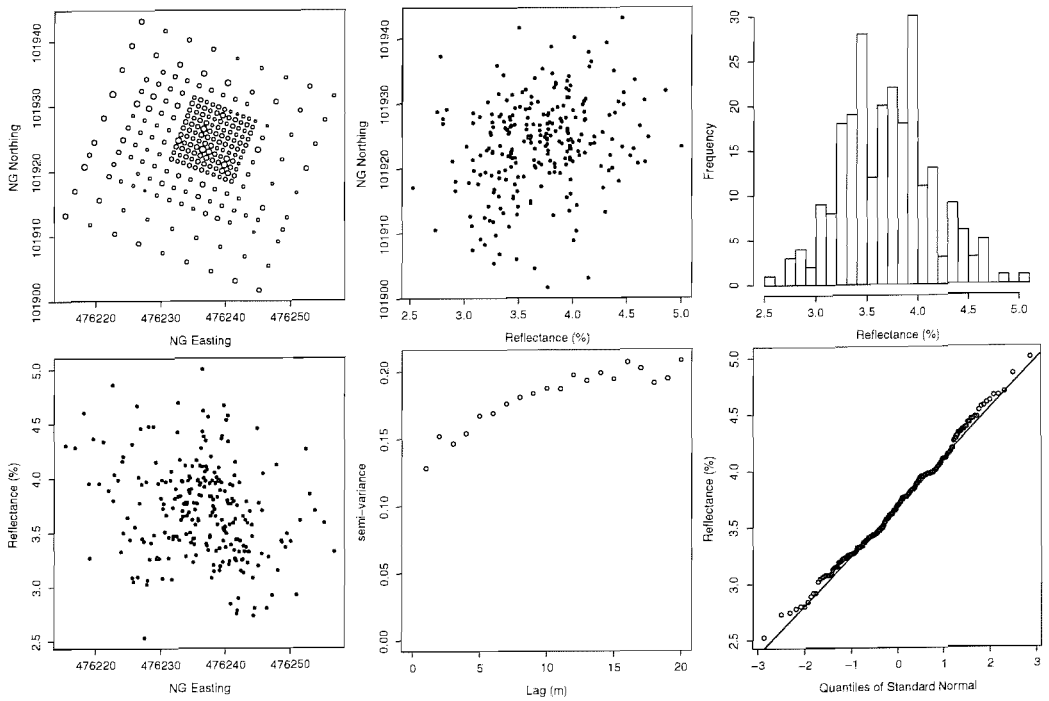


(c) MMR Band 3

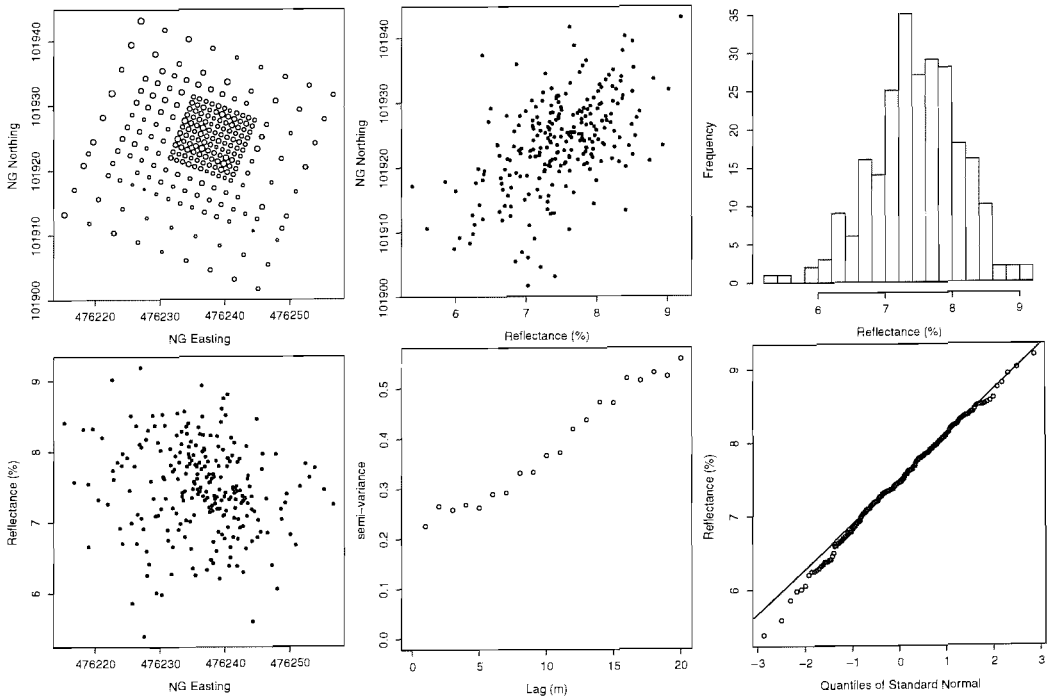


(d) MMR Band 4

FIGURE 4.6: continued

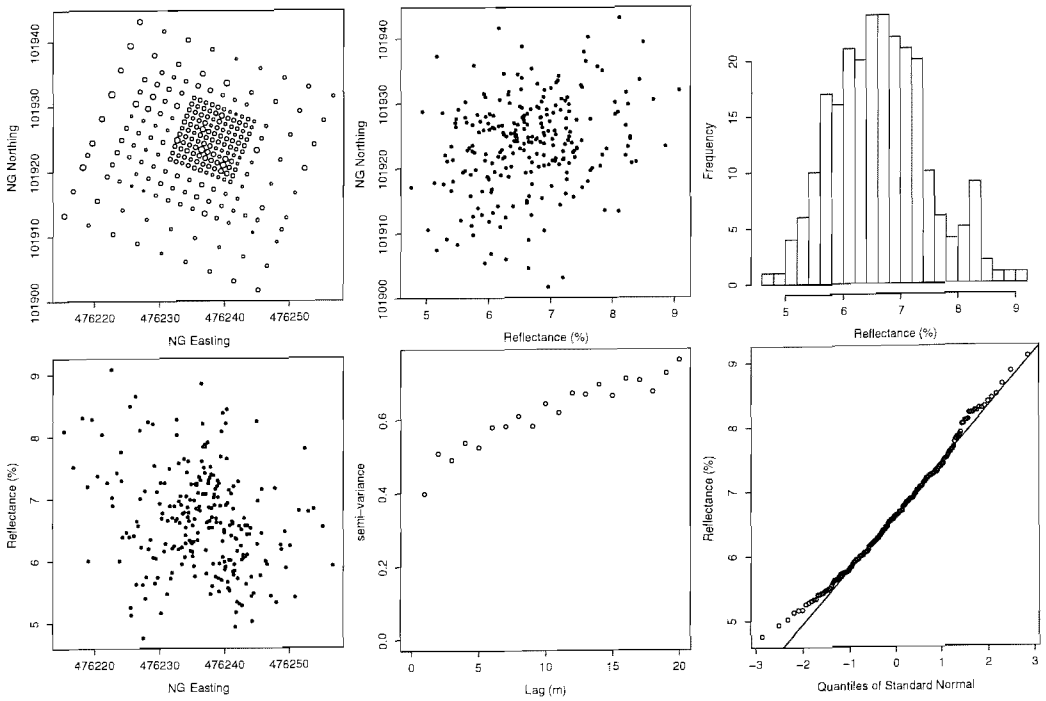


(a) MMR Band 1

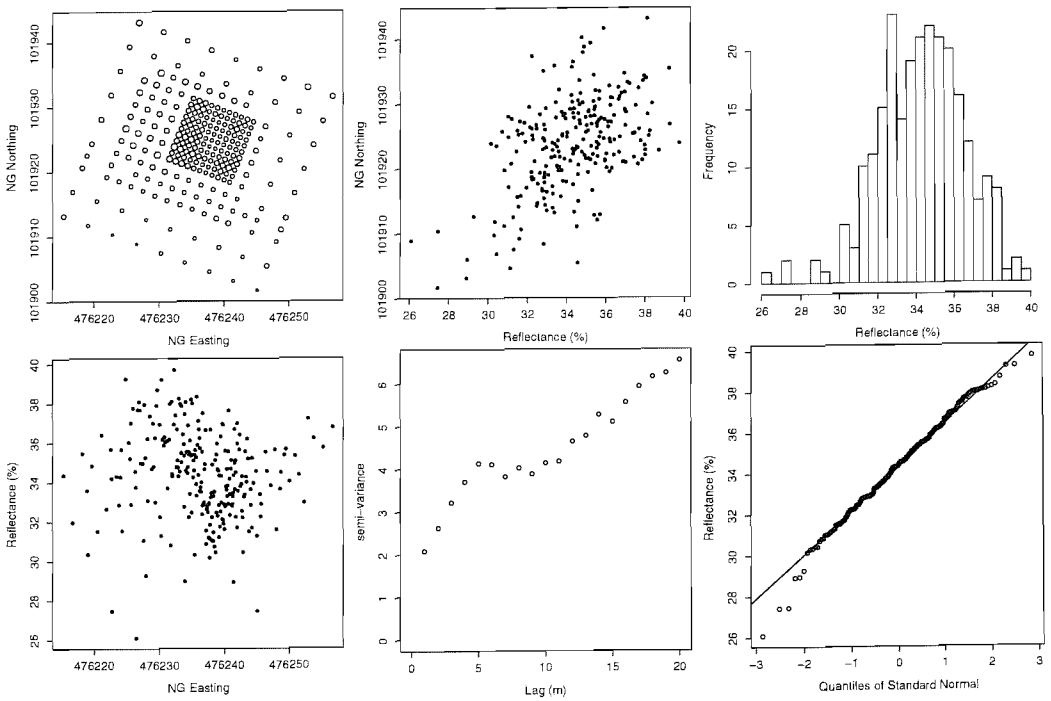


(b) MMR Band 2

FIGURE 4.7: Spatial summary plots for the MMR reflectance measurements taken over the grass surface. Details are given in figure 4.5.



(c) MMR Band 3



(d) MMR Band 4

FIGURE 4.7: continued

4.5.2 Image data

In this section, summary statistics for the image data are presented. Summary plots for the radiance data gathered using the ATM are shown in figures 4.8, 4.9 and 4.10. Summary statistics are given in tables 4.16, 4.17 and 4.18. The format is the same as for the MMR data (section 4.5.1).

It should be recognised that, although these airborne measurements were made over the same GTs as the field targets it should not, necessarily, be expected that they will show the same histograms and QQ plots as the field data. Notably, the measurements are defined on different supports and it is known that measurements of the same phenomenon that are defined on different supports may exhibit different histograms (Armstrong 1998, Bierkens et al. 2000) (see section 2.2.1 for discussion). Furthermore, the measurements were made in different ways. The ATM and MMR have slightly different spectral responses and the two instruments have quite different IFOVs (see section 4.3.1). Note also that the atmosphere affects the remotely sensed measurements. As well as attenuating the signal, as the optical depth and view angle increase, so the amount of radiation reflected from the surroundings surface and atmosphere into the sensor path increases (see section 3.1.1). This has the effect of "blurring" the image and, effectively, reducing the spatial resolution (Kaufman 1989).

For the asphalt surface, figure 4.8 and table 4.16 suggest that the data are normally distributed in each band. However, there is some slight indication of skewness in bands 3, 5 and 7. Skewness is indicated by the concave shape of the data points on the QQ plot (Chambers et al. 1983) and by the skewness index in table 4.16. This effect is most pronounced for band 7. Overall, the shape of the histograms and QQ plots for the ATM bands are slightly different to their comparable MMR bands.

For the concrete surface, figure 4.9 and table 4.17 show that the data are not normally distributed. This is most apparent in bands 2, 3 and 5 but less apparent in band 7. For band 2 there is a noticeable jump towards the right hand side of the QQ plot and the histogram suggests that the distribution is bimodal. For bands 3 and 5 the QQ plots show a concave shape. Together with the histogram this may suggest a skewed, or possibly bimodal, distribution. For band 7 the QQ plot and histogram suggest short tails to the distribution. Overall, the histograms and QQ plots for the concrete surface display similar trends to the comparable bands from the MMR data.

For the grass surface (figure 4.10 and table 4.18) bands 2 and 3 are approximately normally distributed. There is some indication of short tails in bands 3 and 7 and of asymmetry (skewness) in band 5. This is comparable to MMR bands 1 and 3, but quite different to MMR bands 2 and 4. The QQ plots for MMR bands 2 (ATM band 3) and 4 (ATM band 7) show evidence of a skewed distribution. However, the QQ plots and histograms for the comparable ATM bands (3 and 7) clearly indicate short tails.

For the asphalt surface (figure 4.8) the plots show evidence of spatial structure. Pixels

Waveband	Minimum	Median	Mean	Maximum	St. Dev	Skewness	Kurtosis
Band 2	3405	3607	3609	3840	74.63	0.18	0.24
Band 3	3481	3631	3635	3837	61.26	0.34	0.01
Band 5	3295	3451	3453	3613	55.19	0.32	0.22
Band 7	3033	3191	3194	3372	57.60	0.36	0.25

TABLE 4.16: Summary statistics for the ATM data gathered over the asphalt surface. Units are radiance (W m^{-2}).

(represented as points) that are close together appear to have similar values. For bands 2, 3 and 5 there is evidence of a trend increasing from west to east. This is curious, since the trend increases from east to west for the MMR measurements for the comparable MMR bands (Bands 1, 2 and 3). For ATM Band 7 there is no clear trend. The sample variograms for the image data show clear evidence of spatial structure. The broad patterns are similar to the comparable MMR bands although the variogram shapes differ. This can be attributed to three factors: *(i)* the data are measured on different supports, *(ii)* the data are obtained in different ways and *(iii)* sample variograms can be unstable and can be affected by sample size, sample structure, the presence of outliers and the chosen bin and tolerance. These issues are addressed in sections 2.2.1 and 2.2.2.

For the concrete surface (figure 4.9) the plots show evidence of spatial structure. There is evidence of a patch of brighter values in the north–east corner in all bands. Aside from this, there is no evidence of a trend in bands 2 and 3, although there is evidence of a trend (decreasing from west to east) in bands 4, 5 and 7. These broad patterns are similar to those observed for the MMR measurements. The sample variograms show clear evidence of spatial structure. The broad patterns are similar to the comparable MMR bands, although the ATM bands tend to exhibit longer range variation.

For the grass surface (figure 4.10) the plots show evidence of spatial structure. There is evidence of a trend decreasing from west to east in bands 2, 3 and 5. In band 7 there is evidence of a trend decreasing from north to south. These broad patterns are comparable to those observed for the MMR measurements. The sample variograms show clear evidence of spatial structure. For ATM bands 2 and 5 this is more pronounced than the comparable MMR bands (1 and 3).

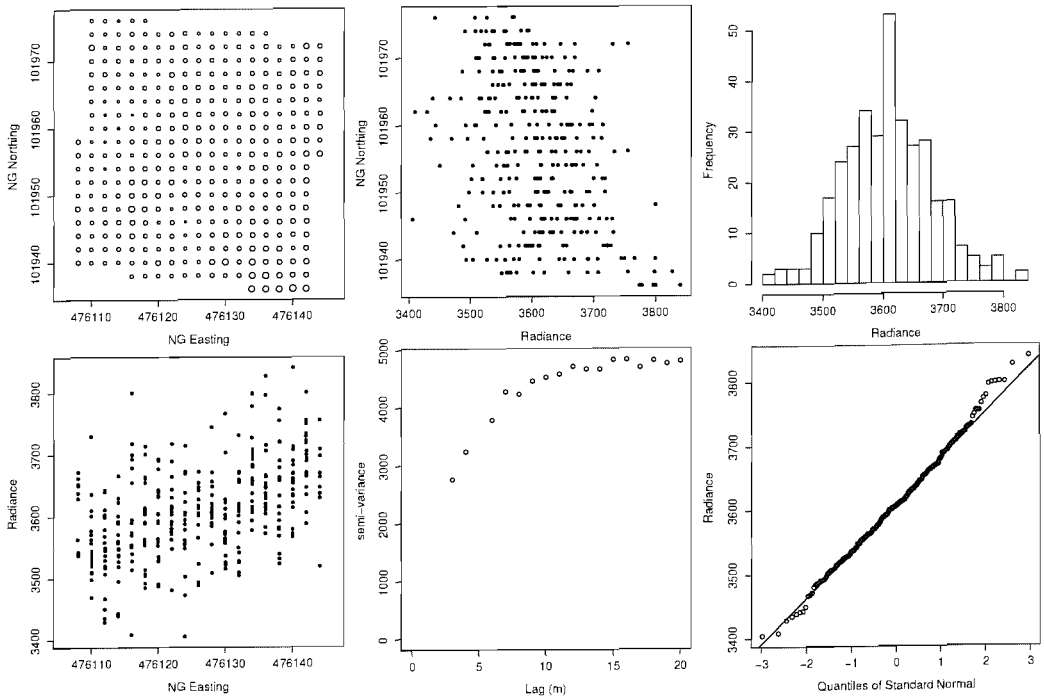
Overall, the above discussion shows that the airborne ATM data show the same general patterns as the MMR data. However, the details differ and the distribution of (pixel-based) radiance measurements made using an airborne sensor (the ATM in this case) cannot be assumed to exhibit the same distributional properties or spatial structure as reflectance measurements made using a field-based radiometer.

Waveband	Minimum	Median	Mean	Maximum	St. Dev	Skewness	Kurtosis
Band 2	4671	5516	5598	7069	404.54	1.42	2.45
Band 3	5784	6858	6915	8360	452.34	0.61	0.35
Band 5	5832	6836	6890	8150	459.78	0.42	-0.17
Band 7	5259	5997	6017	6995	330.35	0.21	-0.53

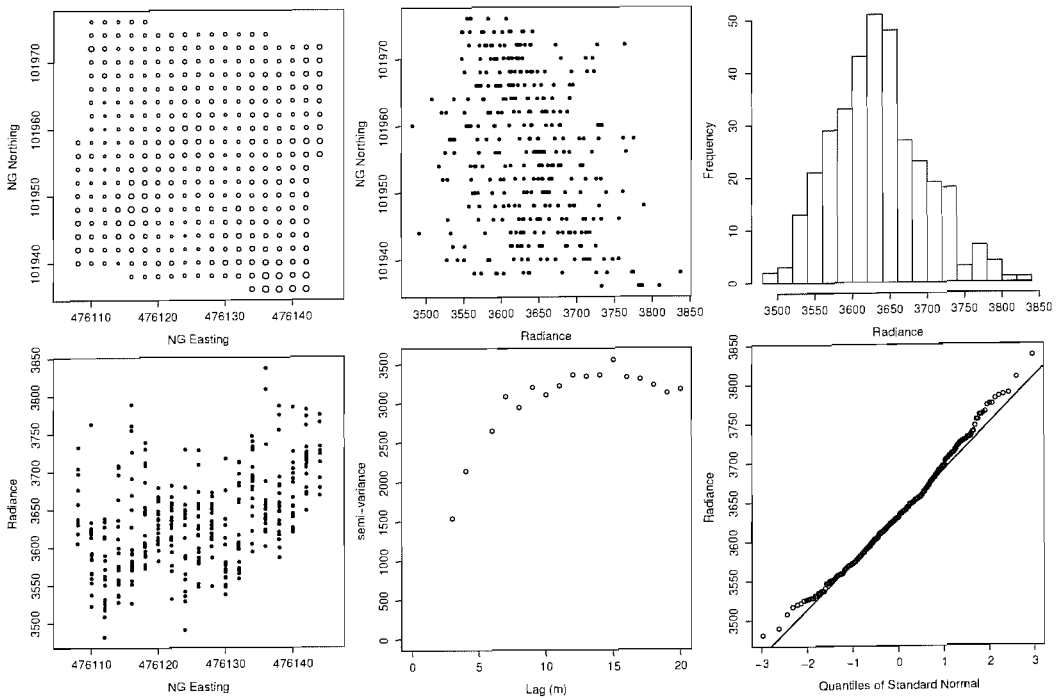
TABLE 4.17: Summary statistics for the ATM data gathered over the concrete surface. Units are radiance (W m^{-2}).

Waveband	Minimum	Median	Mean	Maximum	St. Dev	Skewness	Kurtosis
Band 2	1772	2045	2061	2414	128.81	0.33	-0.33
Band 3	2219	2650	2661	3098	179.77	0.19	-0.49
Band 5	1919	2398	2427	3009	227.73	0.43	-0.38
Band 7	6295	6900	6936	7642	291.70	0.26	-0.79

TABLE 4.18: Summary statistics for the ATM data gathered over the grass surface. Units are radiance (W m^{-2}).

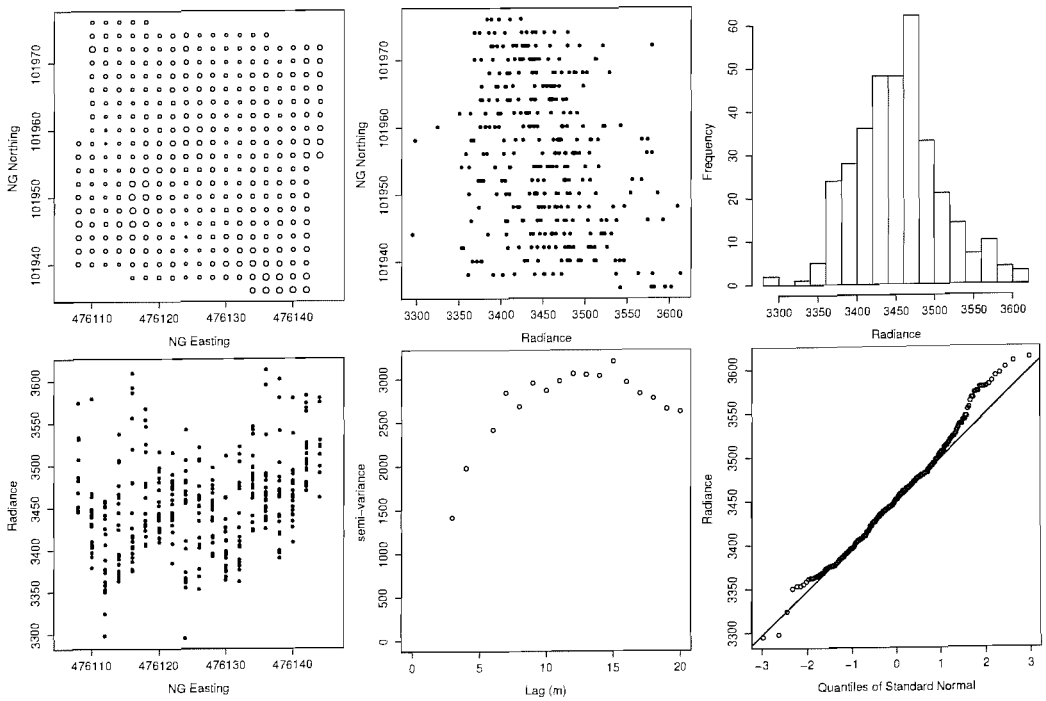


(a) ATM Band 2

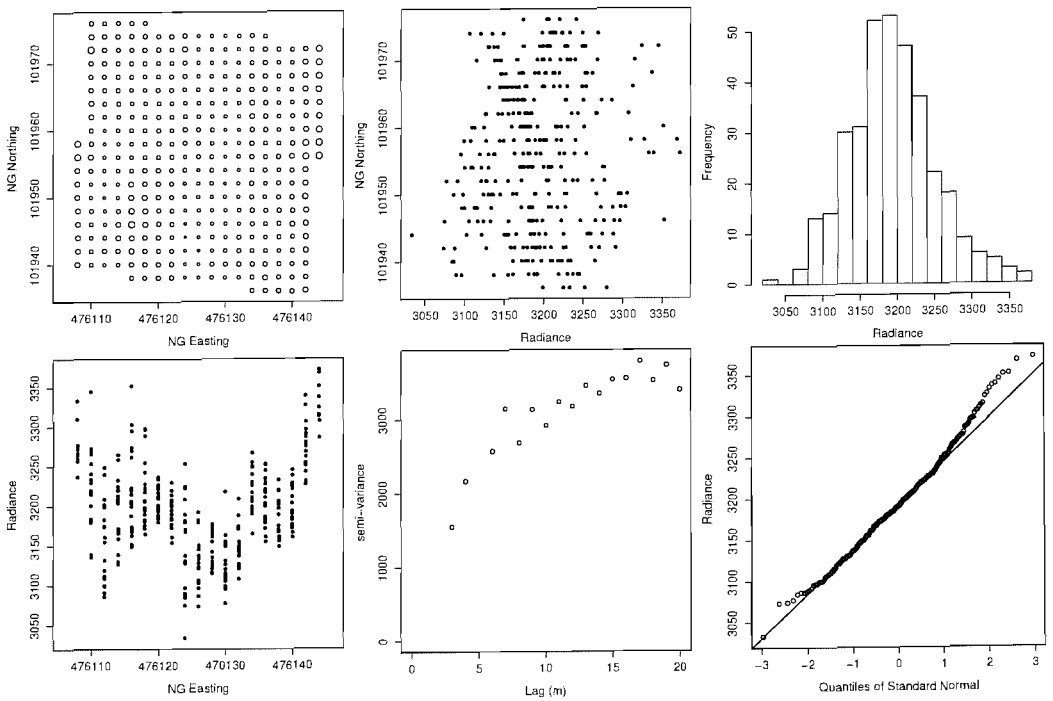


(b) ATM Band 3

FIGURE 4.8: Spatial summary plots for the ATM radiance measurements taken over the asphalt surface. Details are given in figure 4.5.

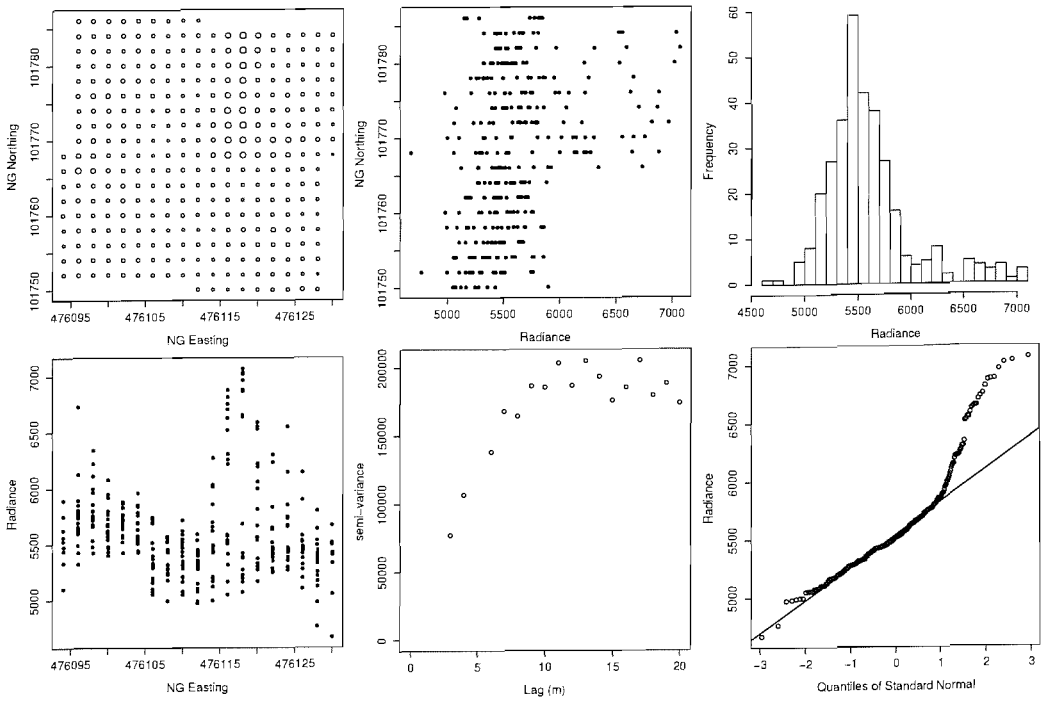


(c) ATM Band 5

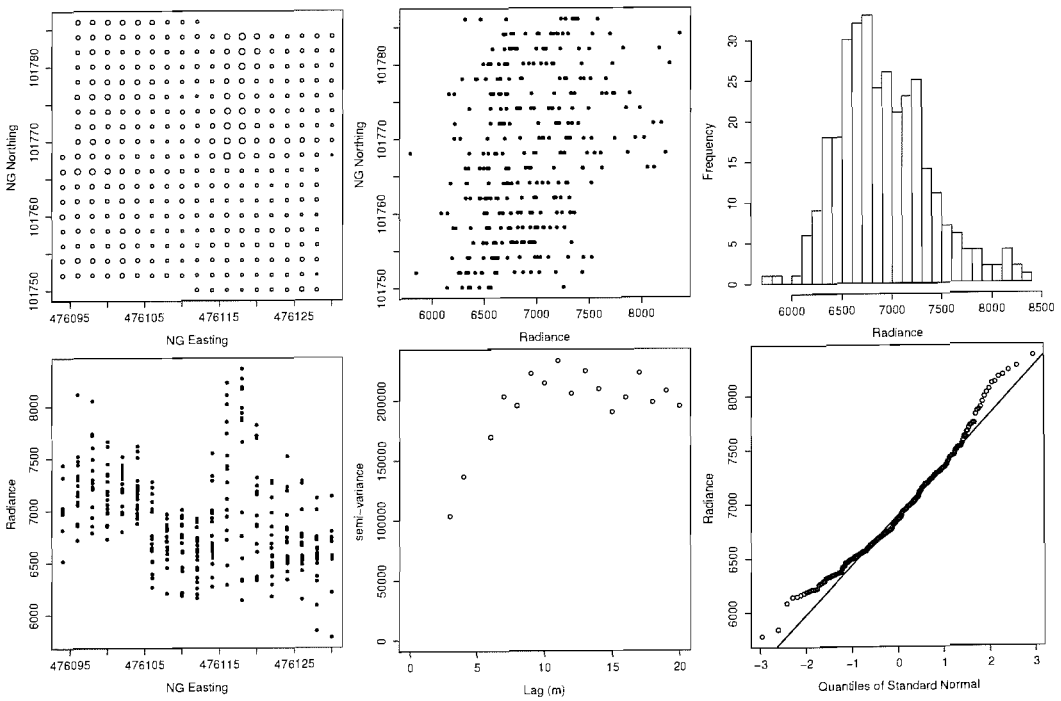


(d) ATM Band 7

FIGURE 4.8: continued

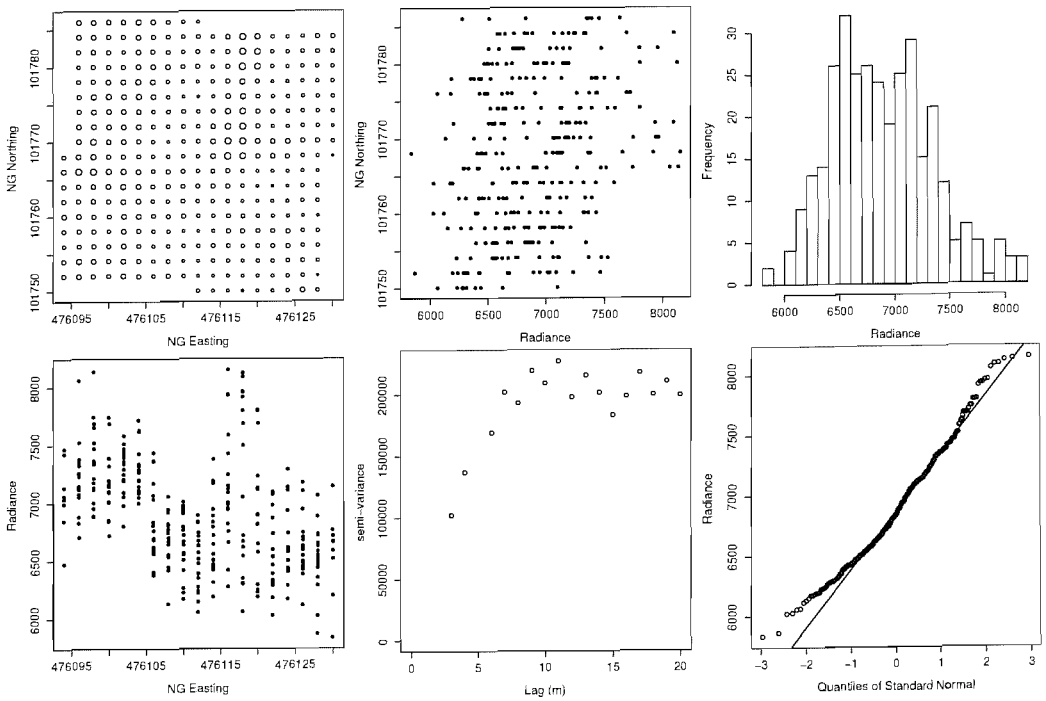


(a) ATM Band 2

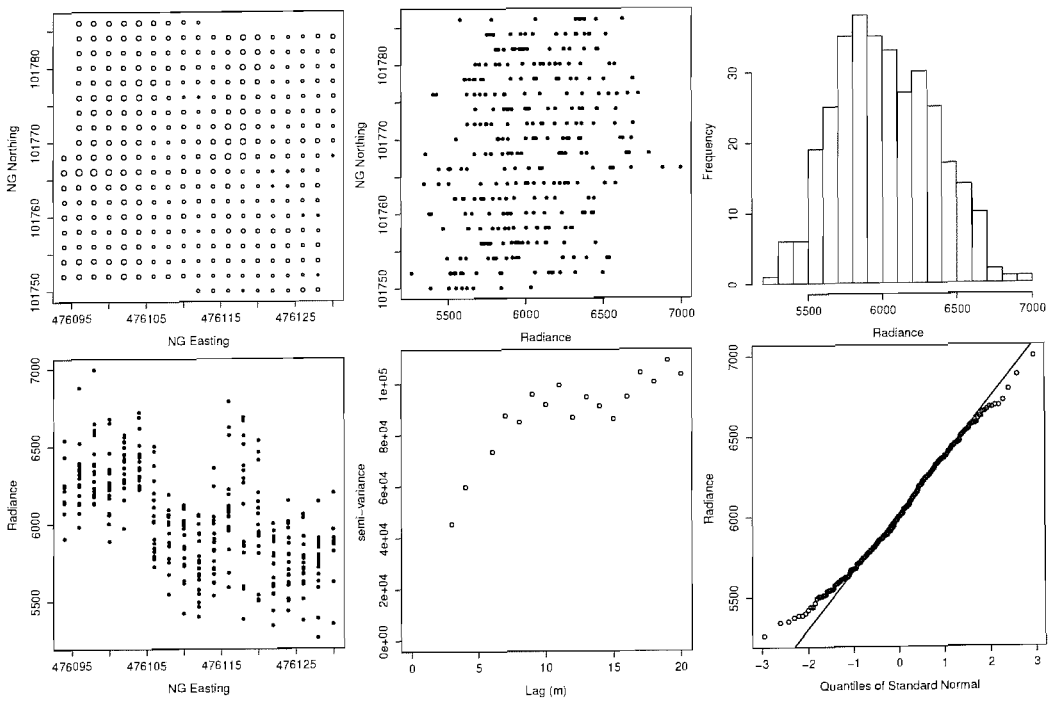


(b) ATM Band 3

FIGURE 4.9: Spatial summary plots for the ATM radiance measurements taken over the concrete surface. Details are given in figure 4.5.

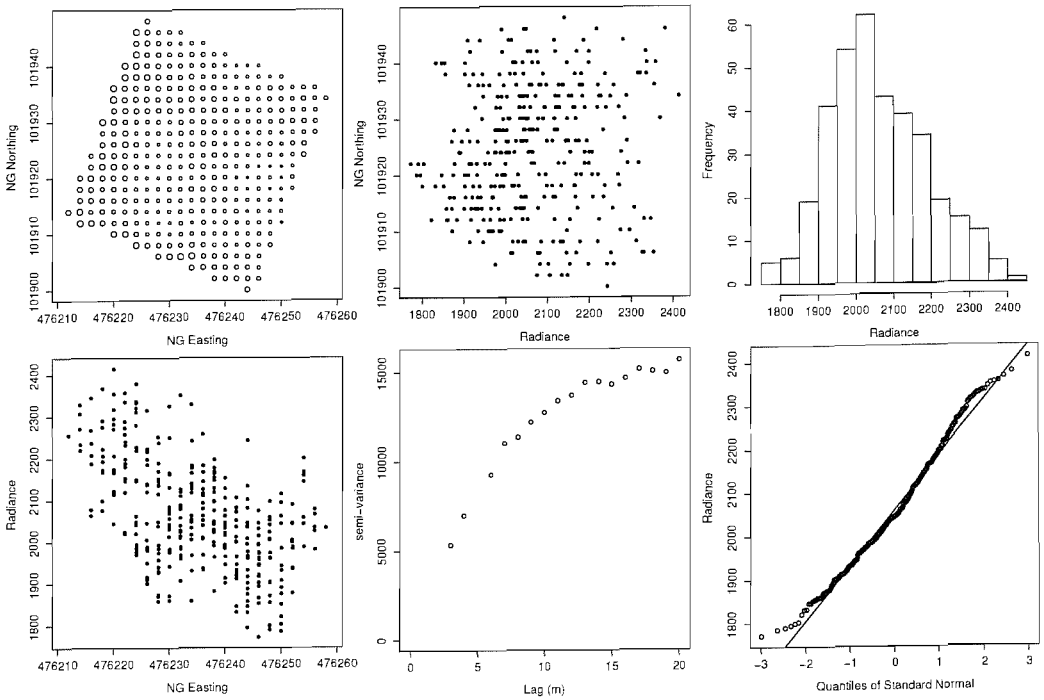


(c) ATM Band 5

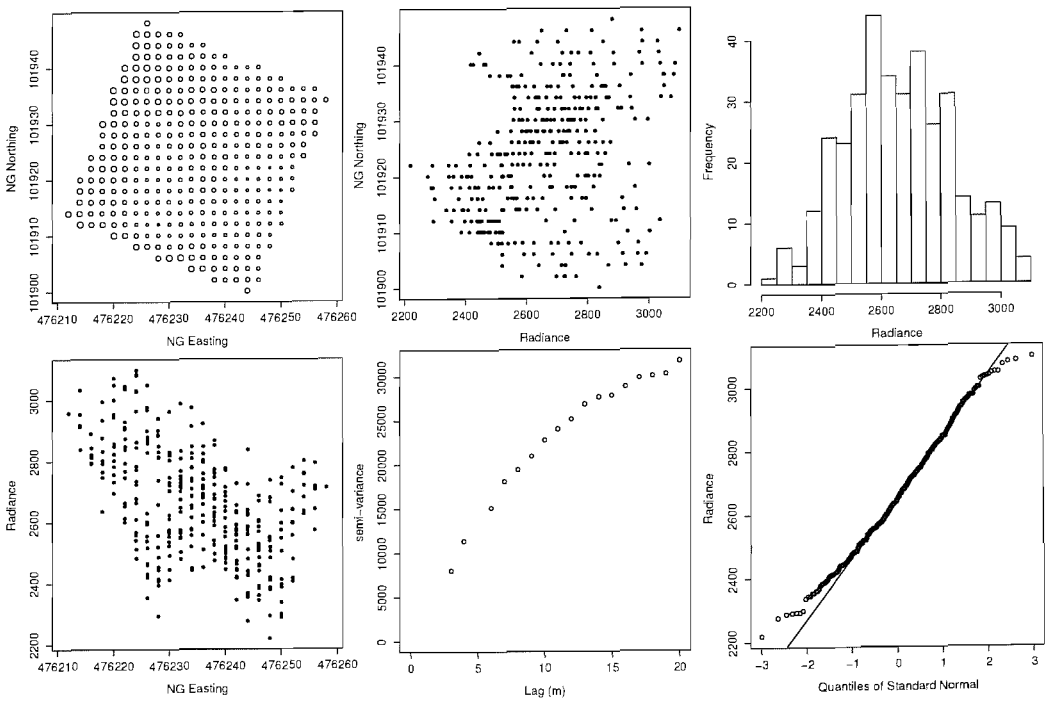


(d) ATM Band 7

FIGURE 4.9: continued

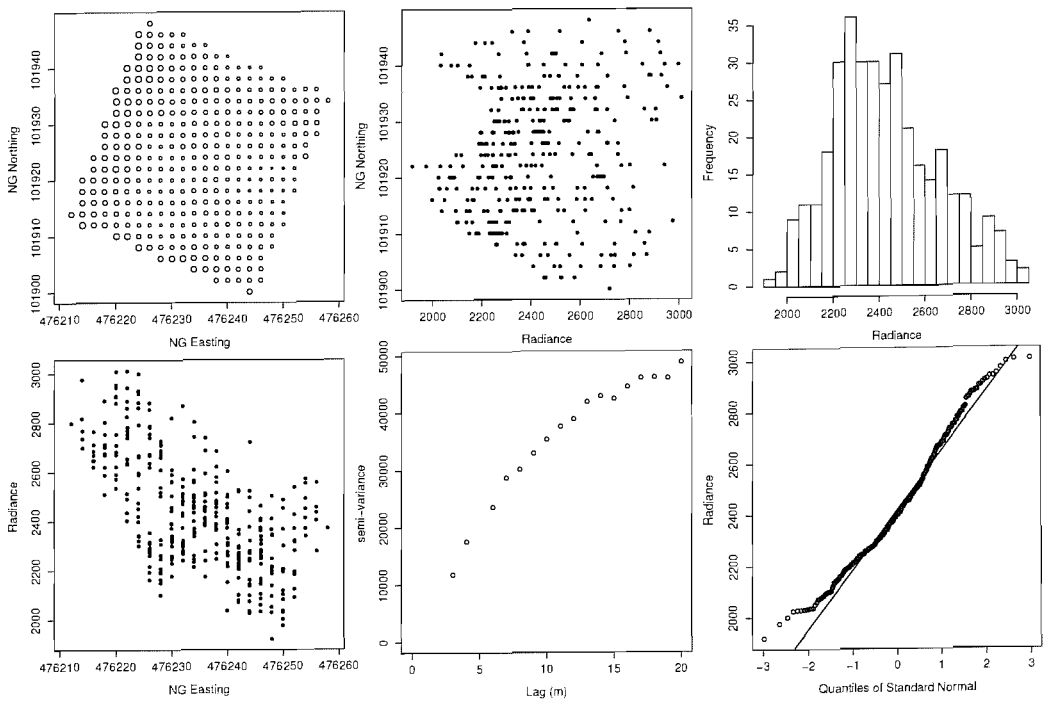


(a) ATM Band 2

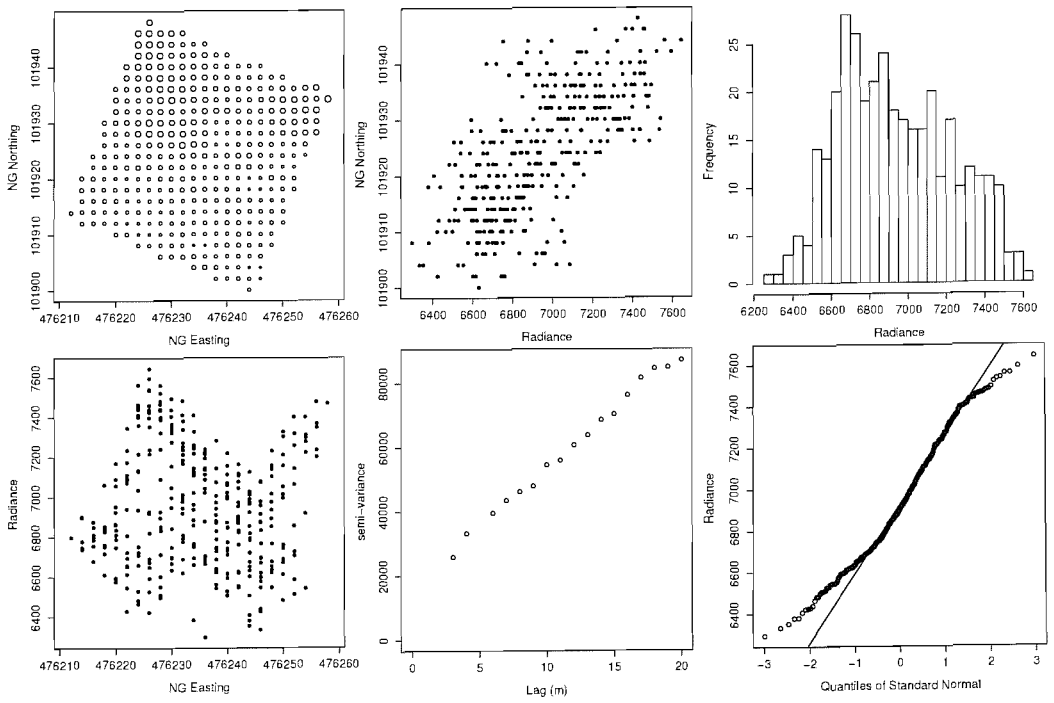


(b) ATM Band 3

FIGURE 4.10: Spatial summary plots for the ATM radiance measurements taken over the grass surface. Details are given in figure 4.5.



(c) ATM Band 5



(d) ATM Band 7

FIGURE 4.10: continued

4.6 Summary

This chapter began by describing the field site at Thorney Island. It then went on to describe the ARSF instruments and MMR together with the pre-processing that was undertaken for this thesis. The field sample for the experimental analysis undertaken in chapters 5, 6 and 7 was outlined. Finally, summaries of the relevant field and remotely sensed data were presented and discussed.

Chapter 5

Analysis I: data pairing

This chapter addresses the issue of data-pairing and gives detailed attention to the mechanisms for combining remotely sensed and field data. This is the first of the key issue for investigation that was outlined in section 1.3. The impact of sample size on the pairing mechanism is also considered. The pairing mechanisms considered are as follows:

- Section 5.1 considers the *"typical"* approach to implementation of the ELM. Essentially pairing is undertaken on a *per target* basis.
- Section 5.3 considers the *"point–pixel"* approach. In this case, each field measurement was paired with the co–located pixel.
- Section 5.4 considers the *"block–pixel"* approach. In this case, the field data were aggregated to pixel-sized support and then paired.

Analysis in this and subsequent chapters was mainly undertaken using the S-Plus[1] and R[2] statistical programming environments. S-Plus and R provide implementations of the S statistical programming language which was initially developed at Bell Laboratories in the 1970s. S-Plus is a commercial implementation, whereas R is a free version available under the GNU public licence. S-Plus has been widely used by statisticians and scientists for several decades owing to its advanced statistical functionality. R grew in popularity over the course of this PhD project and was particularly useful owing to the availability of well-documented third-party libraries, such as `gstat` (Pebesma & Wesseling 1998, Pebesma 2003) and `geoR` (Ribeiro & Diggle 2001, Ribeiro Jr. et al. 2003) for geostatistical analysis. Where greater computational speed was necessary functions were written in C and linked to S-Plus.

[1] For S-Plus see www.splus.com
[2] For R see www.r-project.org

5.1 The typical approach

The *typical approach* is used here to refer to the typical method of combining the field and remotely sensed data in order to estimate the parameters of the regression model. This is the approach most commonly found in the literature (section 3.2.3). The ground targets (GTs) are selected and several field measurements of reflectance are made. The mean and variance of those measurements are then estimated (see tables 4.13, 4.14 and 4.15). The GT is then located in the image and the mean and variance of those pixel-based measurements of radiance is estimated. The mean values of radiance and reflectance are combined and used to estimate the parameters of the regression model. The regression model is given in equation 2.1 (section 2.1), and is repeated below:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{5.1}$$

where y_i is the reflectance measurement, x_i is the radiance measurement and $\varepsilon_i \sim N(0, \sigma^2)$ (σ^2 is the variance).

The typical approach was implemented for the Thorney Island data set. The reflectance in each MMR band was used as the dependent variable and the radiance in the appropriate ATM band was used as the predictor variable. Since there are three GTs (asphalt, concrete and grass) this gave three points for calculating the parameters (β_0, β_1 and σ^2) of the regression model for each waveband. The results of the parameter estimation are given in table 5.1. Plots showing the fitted regression line and 95% confidence and prediction intervals are shown in figure 5.1. Note that the confidence interval reflects uncertainty in the location of the regression line whereas the prediction interval reflects the uncertainty in prediction. The prediction interval will always be wider than the confidence interval and may be substantially wider, depending on the magnitude of σ^2. The regression parameters were calculated using ordinary least squares (OLS). If a Bayesian framework is applied and non–informative priors adopted, the same results are found.

The p values given in table 5.1 were used to test the null hypothesis that $\beta_0 = 0$ or $\beta_1 = 0$ by applying a t–test (Neter et al. 1996). This showed that, for many cases, the estimated parameters were not significantly different to zero. Clearly, the decision to choose the 95% or 99% confidence level is important here. However, given the importance of atmospheric correction in remote sensing applications, it is reasonable to suggest that the user would be demanding and choose a high confidence level (99% or higher). Given these results it might be argued that there is no valid regression and, therefore, no basis on which to apply the ELM. This interpretation would be reasonable if there was no further information available. However, in this case, it is misleading because there is prior knowledge about the form of the relationship. It was demonstrated in section 3.2.3 that there is a physical basis for this relationship. Notably it is expected that $\beta_1 > 0$ and $\beta_0 < 0$. The regressions shown in figure 5.1, together with 95% confidence and prediction intervals, all indicate a trend that, broadly, conforms to expectation. However, these plots

also emphasise that, on the basis of the data used, there is a great deal of uncertainty about exactly where the regression line should lie.

In all cases, the R^2 value was very large ($R^2 > 0.99$). Hence, this analysis demonstrated why the R^2 should not be used as the basis for assessing the accuracy of a regression model. It is tempting to take this large R^2 as an indication that the model is "good", although, in this case, that interpretation would be incorrect.

The plots (figure 5.1) show the best fit regression line and the 95% prediction interval. The 95% prediction interval should be interpreted as the 95% prediction interval for any given level of x. It is *not* the simultaneous prediction interval for all possible x levels. Recall from section 2.1 that the prediction error for a new observation x_{new} is:

$$\begin{aligned} Var(y_{new}) &= Var(\varepsilon) + Var(\hat{y}_{new}) \\ &= \sigma^2 + \sigma^2 \left(\frac{1}{n} + \frac{(x_{new} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right) \end{aligned} \tag{5.2}$$

where $Var(\varepsilon) = \sigma^2$ and $Var(\hat{y}_{new})$ is the variance of the sampling distribution of $\hat{y}_{new}$. Hence, it is clear that the variance of the new prediction is strongly dependent on:

1. the estimated value of σ^2;
2. the number of data points, n;
3. $x_{new} - \bar{x}$.

Under OLS, the estimate of σ^2 is given by the MSE (mean square error) as:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-2} = \frac{SSE}{n-2} \tag{5.3}$$

where y is the dependent variable, n is the number of samples and SSE is the sum of square error. Note that $n - 2$ gives the *degrees of freedom*, which is reduced by 2 owing to the need to estimate β_0 and β_1. Hence, for three GTs ($n = 3$), there is only one degree of freedom. This leads to uncertainty in the estimation of σ^2, an issue which is also reflected in the associated t-value. The t-value is important since it is used for hypothesis testing on the parameter estimates and for assigning confidence and prediction intervals (see section 2.1). Tables of the percentiles of the t-distribution (e.g., Neter et al. 1996) show that, for low degrees of freedom, the t-value increases sharply.

It is clear that, apart from MMR bands 1 and 3 (figures 5.1(a) and 5.1(c)), the 95% confidence and prediction intervals were very wide. Even for bands 1 and 3, the confidence interval of the regression line was *at least* 2% reflectance. This is attributed to the magnitude of $\hat{\sigma}^2$ and the number of data points (i.e., $n = 3$). This means that if the points deviate only slightly from a straight line then the prediction interval increases greatly. This problem is inherent where regression is conducted with such small samples.

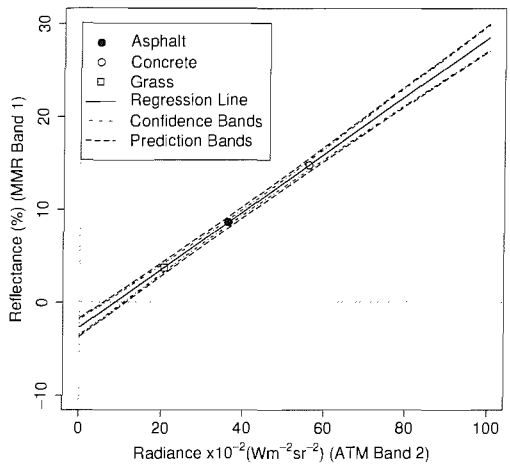
	MMR B1 on ATM B2	MMR B2 on ATM B3	MMR B3 on ATM B5	MMR B4 on ATM B7
$\hat{\beta}_0$	-2.7061	-0.7807	-1.9450	-4.1020
$s.e.(\hat{\beta}_0)$	0.0661	0.7683	0.0029	2.0472
p	0.0156	0.4949	0.0251	0.2947
$p < 0.05$	Yes	No	Yes	No
$p < 0.01$	No	No	No	No
$\hat{\beta}_1$	0.3116	0.2985	0.3553	0.5453
$s.e.(\hat{\beta}_1)$	0.0016	0.0161	0.0016	0.0365
p	0.0034	0.0344	0.0029	0.0425
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	No	Yes	No
$\hat{\sigma}^2$	0.0017	0.2583	0.0029	1.0115
R^2	> 0.99	> 0.99	> 0.99	> 0.99

TABLE 5.1: Table showing the result of parameter estimation using of the ELM regression model using the "typical approach". The parameters were estimated using classical least squares and implemented in S-Plus.

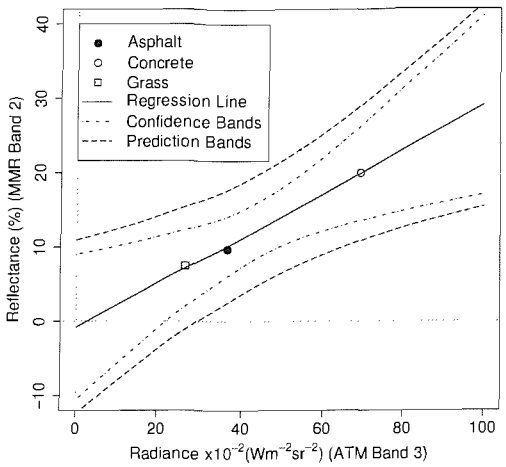
The above analysis has revealed two key problems with the typical implementation of the ELM, as follows:

- By averaging over GTs the within-target variability is lost. This is acceptable if the objective is to then apply the ELM over GT sized supports. However, the objective is to apply the ELM over pixel sized supports. Hence, the typical approach is likely to underestimate σ^2.
- Averaging over GTs also has the effect of reducing the number of data pairs in the regression model such that n equals the number of GTs (three in this case). This means that there are very few degrees of freedom – leading to uncertainty in the estimate of σ^2 and inflation of the t-value. This potentially leads to very wide confidence and prediction intervals and even to the situation where the null hypothesis (i.e., that $\beta_0 = 0$ and $\beta_1 = 0$) cannot be rejected. Clearly, this latter conclusion is incorrect because it is known from theory that $\beta_0 < 0$ and $\beta_1 > 0$. This is interpreted as evidence of large uncertainty in the estimation of the regression parameters and, hence, the location of the regression line.

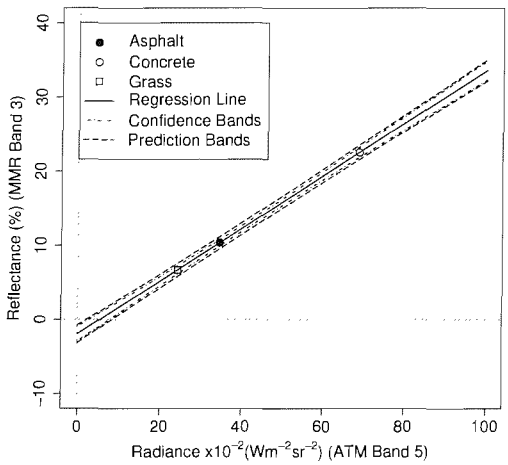
It might be argued that the objective of averaging over targets is to provide an accurate and stable characterisation of the target as a whole and that reducing variability is sensible because the objective is to determine accurately the empirical line rather than to determine accurately prediction intervals (Milton et al. 2000). From one perspective, this is reasonable since the ELM constitutes a form of calibration hence the primary interest lies in the location of the regression line and the confidence bands around it (Webster 1989). Hence, in order to explore the impact of uncertainty in the ELM we might vary the location of the regression line within the confidence intervals. However, if the objective was to predict a biophysical variable then primary interest would be in the prediction intervals.



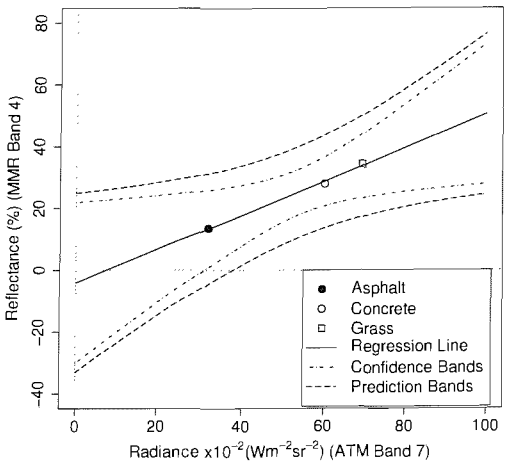
(a) ELM for MMR Band 1



(b) ELM for MMR Band 2



(c) ELM for MMR Band 3



(d) ELM for MMR Band 4

FIGURE 5.1: ELM regressions for at-surface reflectance (MMR) against at-sensor radiance (ATM). For each graph the regression line and 95% confidence and prediction intervals are shown.

It might be further argued that the uncertainty in the location of the regression line, given by the confidence intervals in figure 5.1, is *over estimated*. The basis for this argument is that the value of each data point has been accurately fixed by averaging over a large number of measurements for each GT. Since there is a high degree of confidence in each GT it might be argued that the uncertainty in the regression line should be low. Hence, it might be argued that we should look for a mechanism for "adding back" the information lost by averaging – perhaps by weighting each point according to the variability of the target.

The problem with the above arguments is that accurate determination of both the regression line (and its associated confidence intervals) *and* the prediction intervals are contingent on σ^2 and the number of degrees of freedom. Hence, if the third data point

departs from a straight line the confidence and prediction intervals are, quite rightly, inflated. This recognises uncertainty in the location of the regression line. Indeed, in the scatter plots presented by Karpouzli & Malthus (2003), most of the nine GTs used do *not* lie on the ELM regression line. This illustrates that the ability to accurately determine reflectance of a GT does not mean that uncertainty in the location of the regression line should be low, particularly when n is small. Furthermore, applying weighted least squares (WLS) may even lead to a further increase in the width of the confidence interval, since the concrete target is downweighted owing to its relatively large variability. Hence, it is argued here that the confidence intervals in figure 5.1 quite rightly illustrate uncertainty in the ELM caused by using a small number of GTs. Hence, rather than trying adjust the confidence intervals, sections 5.3 and 5.4 address this by implementing the regression using points and pixels respectively. These allow the data to be fully used in the ELM.

Finally, note that the above implementation of the typical approach utilised a large data set (approximately 200 field-measurements per target). Typically far few measurements are taken. The next section (section 5.1.1) considers the implication of sample size.

5.1.1 The implications of sample size

The analysis reported above (section 5.1) demonstrated the implementation of the typical approach for the full data set collected for this project. This comprised more than 190 field measurements per GT. In practice, far fewer measurements are taken (see section 3.2.3.4 for further discussion). Smith & Milton (1999) take between 5 and 35 samples per GT and Karpouzli & Malthus (2003) take between 15 and 60 measurements per target, although Clark et al. (2000) and Moran et al. (2001) invest very substantial time in characterising GTs. Therefore, it is important to consider the implications of reducing the sample size on the implementation of the typical approach. In particular, the following issues are considered:

1. To what extent does reducing the sample size increase the width of the confidence interval on the parameter estimates and on the regression line? In the extreme case, this will be exemplified when the t–test shows that the estimated parameter is not significantly different to zero. As noted above, this should not lead to the conclusion that there is no relationship be radiance and reflectance but can be interpreted as evidence of large uncertainty in the location of the regression line.

2. To what extent does reducing the sample size increase the variability of the parameter estimates?

3. To what extent does reducing the sample size increase the size of $\hat{\sigma}^2$?

All of these issues are important since they lead to uncertainty when parameterising and implementing the ELM. They are also particularly important for the GTs used in this

Sample size	Sample mean	Sample variance
1	-2.48050	NA
3	-0.48655	0.38355
5	-0.43504	0.51722
10	-0.04142	1.58996
20	0.56835	1.01793
30	0.01672	0.62516
50	0.20761	0.92002
100	0.14061	1.27643
200	0.04638	1.04430
1000	0.00168	0.97527

TABLE 5.2: Table illustrating the estimated parameters (mean and variance) for samples drawn from a standard normal distribution ($N(0,1)$). One realisation of a data set is taken per sample size.

thesis since it was shown that all targets exhibit spatial structure. Furthermore, the reflectance measurements obtained over the concrete (and to a lesser extent the grass) showed evidence of being skewed.

The sources of uncertainty described above arise because the y_i in equation 5.1 are calculated as an average of a sample. This sample average is an estimate of the population average and may deviate from it, particularly for small samples. This has implications for the ELM, where the y_i are based on the average of a small sample. This effect is illustrated in table 5.2 which shows one realisation of a sample from a standard normal distribution (i.e. $N(0,1)$) and the associated *sample* mean and variance.

To address the above issues, an experiment was designed in which sub–samples of the measurements for each GT were taken. The sample sizes were the full data set and then 100, 50, 30, 20, 10, 5, 3 and 1 measurements per GT. These were labelled $nFull$, $n100$, $n50$, $n30$, $n20$, $n10$, $n5$, $n3$ and $n1$ and the sample was set up such that $n1 \subset n3 \subset n5 \subset n10 \subset n20 \subset n30 \subset n50 \subset n100 \subset nFull$. Each subset was drawn at random from the next largest subset in the chain, allowing comparison from one sample size to the next. One thousand such samples were drawn, obtaining $n_1100, n_2100, \ldots, n_{1000}100$ and $n_150, n_250, \ldots, n_{1000}50$ and so on.

Table 5.3 shows the proportions of the simulations where $\beta_0 \neq 0$ or $\beta_1 \neq 0$ for both the 95% and 99% confidence level. For MMR band 1 (table 5.7(a)), it was clear that, as the sample size decreases, the proportion of cases where the parameter does not equal zero also decreased. This was least apparent for β_1 at $p = 0.05$ where for $n10$ and above over 99% of the realisations found that $\beta_1 \neq 0$. However, as mentioned above, the user is likely to require a high degree of confidence in the parameter estimation. For β_1 at $p = 0.01$ the proportion of cases where β_1 was not significantly different to zero falls off at $n100$. It is interesting to note that for β_0 at $p = 0.01$, β_0 was not significantly different to zero for the full data set, whereas instances where $\beta_0 \neq 0$ were found for $n1$ through to $n100$. This was, most likely, due to the effect illustrated in table 5.2 and that chance provides a

sample for each GT, whose means lie closer to a straight line than the mean for the full data set. A similar effect was observed in other cases. As with the previous section, these results are noteworthy because they reflect uncertainty in the estimation of the regression parameters and *not just* because they lead to rejection of the form of the relationship.

For MMR band 2 (table 5.7(b)) the proportion of cases where β_0 was significantly different to zero was never greater than 8% (for $p = 0.05$ and $p = 0.01$). For β_1 at $p = 0.05$ the proportion of cases where β_1 was significantly different to zero tailed off below $n30$ and was never more than 15% at $p = 0.01$.

For MMR band 3 (table 5.3(c) the proportion of cases where β_0 was significantly different to zero tailed off from $n100$ for $p = 0.05$ and was never greater than 19% for $p = 0.01$. The proportion of cases where β_1 was significantly different to zero tailed off below $n3$ at $p = 0.05$ and from $n100$ for $p = 0.01$.

For MMR band 4 (table 5.7(d) the proportion of cases where β_0 was significantly different to zero was never greater than 9% (for $p = 0.05$ and $p = 0.01$). For β_1 at $p = 0.05$ the proportion tailed off below $n100$. For β_1 at $p = 0.01$, the proportion of cases where β_1 was significantly different to zero was never more than 12%.

The conclusion to the above analysis is that reducing the sample size increases the uncertainty with which one can estimate $\beta_0 \neq 0$ at both $p = 0.05$ and $p = 0.01$. If $p = 0.05$ is used as the criterion for determining whether $\beta_1 \neq 0$ then *30 measurements* are required per target in order to ensure that more than 99% of the realisations give $\beta_1 \neq 0$. However, this will only apply for MMR bands 1 to 3 and *not* to MMR band 4. *If the more stringent criteria of* $p = 0.01$ *is used for determining whether* $\beta_1 \neq 0$ *then none of the sub samples will suffice in any of the MMR bands.*

(a) Parameter estimation for y = MMR Band 1 and x = ATM Band 2.

	$p < 0.05$		$p < 0.01$	
Sample size	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$
$nFull$	100	100	0	100
$n100$	91.9	100	28.1	90.8
$n50$	78.2	100	19.1	74.9
$n30$	64.4	100	14.7	61.5
$n20$	57.6	100	12.5	53.0
$n10$	39.2	99.3	8.9	37.2
$n5$	29.2	94.9	5.6	26.9
$n3$	22.9	83.6	4.4	20.5
$n1$	11.6	55.9	2.8	11.5

(b) Parameter estimation for y = MMR Band 2 and x = ATM Band 3.

	$p < 0.05$		$p < 0.01$	
Sample size	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$
$nFull$	0	100	0	0
$n100$	0	100	0	0
$n50$	0	100	0	0
$n30$	0	99.5	0	0
$n20$	0	98.4	0	0.1
$n10$	0.5	88.9	0.1	2.1
$n5$	1.1	78.8	0.1	4.4
$n3$	2.5	67.0	0.4	7.1
$n1$	7.4	61.2	1.5	14.1

TABLE 5.3: Table showing the proportion of the parameter estimates that are significantly different to zero for each sample size (1000 realisations per sample size). The t-tests are evaluated at the 95% (p=0.05) and 99% (p=0.01) confidence levels.

(c) Parameter estimation for y = MMR Band 3 and x = ATM Band 5.

	$p < 0.05$		$p < 0.01$	
Sample size	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$
nFull	100	100	0	100
*n*100	83.2	100	17.1	99.1
*n*50	69.2	100	18.1	92.3
*n*30	57.2	100	12.6	84.3
*n*20	48.1	100	12.1	75.5
*n*10	36.4	100	7.0	59.8
*n*5	24.8	99.6	5.3	43.4
*n*3	23.1	97.2	5.3	36.4
*n*1	15.4	81.7	3.2	22.6

(d) Parameter estimation for y = MMR Band 4 and x = ATM Band 7.

	$p < 0.05$		$p < 0.01$	
Sample size	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$
nFull	0	100	0	0
*n*100	0	84.6	0	0
*n*50	0.1	73.1	0	0.3
*n*30	0.4	62.1	0.1	0.6
*n*20	2.2	59.7	0.2	3.3
*n*10	5.2	57.1	1.1	8.1
*n*5	8.2	53.7	1.9	11.5
*n*3	7.1	53.0	1.2	11.1
*n*1	3.8	33.6	0.7	7.2

TABLE 5.3: continued

The above analysis illustrated that decreasing the sample size further reduces the confidence in the estimation of the regression parameters. The impact of this variability is explored further below.

The range of variability in the parameter estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$, gained from drawing 1000 realisations for a given sample size, are illustrated using histograms in figures 5.2 and 5.3 respectively. This analysis was conducted for MMR band 1 (paired with ATM band 2). It is clear that the variablity in the parameter estimates increased as the number of measurements per target decreased. The effect of this is illustrated in figure 5.4 . This shows the regression line for the full data set (red line) and the regression line calculated on the basis of ten realisations (black line). The maximum and minimum predicted values at $x = 40$ and $x = 50$ (typical of urban surfaces) are shown in table 5.4. It is clear that the range of variability increases as the sample size increases and the distance from $\bar{x}$ increases. Hence, if the user could tolerate a range of 1% (reflectance), then they would require a sample size of 20 measurements per target at $x = 40$ and 100 measurements per target at $x = 50$. It should be emphasised that this range of values is for the predicted regression line – it does not include the uncertainty given by these confidence and prediction intervals. Hence, although this provides a straight forward visualisation of how variability increases as sample size decreases, the actual uncertainty in estimation and prediction will be larger. This interpretation is predicated on the user accepting $p = 0.05$ for accepting the validity of the regression line. If they set the more stringent criteria of $p = 0.01$, then 100 measurements would be required at both x–levels.

Finally, figure 5.5 shows the histograms of the predicted values of σ^2 for the different sample sizes ($n100$ to $n1$). It should be recognised that $\hat{\sigma}^2 \sim \chi^2_{n-2}$ (a chi-square distribution with $n-2$ degrees of freedom) for all sample sizes (Casella & Berger 1990, Neter et al. 1996). Hence, for $n = 3$ targets $\hat{\sigma}^2 \sim \chi^2_1$. The shape of the histograms were similar at all sample sizes, as expected. However the range of the estimated values of σ^2 increased as the sample size decreased. For $n100$ 60% of the estimates are greater than 0.0017 (i.e. $\hat{\sigma}^2$ for $nFull$), this increases to 90% for $n5$ and 95% for n1.

The magnitude of $\hat{\sigma}^2$ is important since this parameter is used to calculate the confidence intervals on the parameters and prediction intervals for predicting a new value or set of values. It is also required to calculate the t–statistics which shows whether a parameter is significantly different to zero. Hence the increase in the estimate of σ^2 with decreasing sample size will lead to a widening of the confidence and prediction intervals.

5.1.2 The typical approach: conclusions and recommendations

The above analysis and discussion began by examining the regression models for all four MMR bands for the full data set. It was illustrated that, in all cases, the confidence intervals on the regression line were broad. This is illustrative of uncertainty in location of the regression line. For all bands, this led to the situation where a t–test showed that

Sample Size	Predicted Value ($x = 40$)		Predicted Value ($x = 50$)	
	Minimum	Maximum	Minimum	Maximum
$n100$	9.6	10.0	12.5	13.2
$n50$	9.4	10.1	12.3	13.5
$n30$	9.4	10.2	12.2	13.6
$n20$	9.3	10.3	12.1	13.8
$n10$	9.1	10.6	11.8	14.3
$n5$	9.0	11.1	11.7	15.2
$n3$	8.7	11.4	11.3	15.8
$n1$	8.4	12.0	10.9	16.5

TABLE 5.4: Table showing the range of predicted regression values for $x = 40$ and $x = 50 \times 10^{-2}\mathrm{Wm}^{-2}\mathrm{sr}^{-2}$.

one or both of the parameters was not significantly different from zero at $p = 0.01$. It was argued that this should not lead to rejection of the *form* of the relationship but that the data cannot be used for accurate parameter estimation.

It was noted that the number of samples collected for this study was large (at least 190 per target) and that far fewer measurements would typically be taken. It might be argued that this analysis is spurious, since it was already shown that the model is not adequate for the full data set. However, the analysis was useful because it illustrated the potential traps that may arise if a user was to use three targets and a relatively small sample size. The key message here is that, as the sample size decreases the confidence in the regression further decreases. This may also mean that the user finds that the t–tests show that the parameters are not significantly different to zero. The analysis also demonstrated that, as the sample size decreased the variability in the parameter estimates, and hence the location of the regression line increased. For MMR Band 1, regressed on ATM band 2, this is particularly noticeable for less than 50 measurements per target. This was illustrated for MMR band 1 by the variation in the estimated regression line.

The above analysis strongly suggests that the typical approach is likely to be problematic when implemented within a conventional regression framework. This is because of the low confidence in the parameter estimates, even for the full sample. Furthermore, the variability in the parameter estimates and hence the variability in ELM prediction increases as the sample size decreases. However, the form of the relationship is known *a priori* hence the implication is that the conventional regression approach has underestimated confidence in the ELM regression.

It might be argued that concerns over whether $\beta_1 \neq 0$ and over the magnitude of $\hat{\sigma}^2$ and the uncertainty in new variable prediction arise partly because the number of GTs (and hence the number of degrees of freedom) is small. Sampling additional targets would increase n in equations 5.2 and 5.3 and decrease the critical value for the t–test. This latter approach may be regarded as desirable by some users. For example, Karpouzli & Malthus (2003) use up to nine GTs for an implementation of the ELM. Nevertheless,

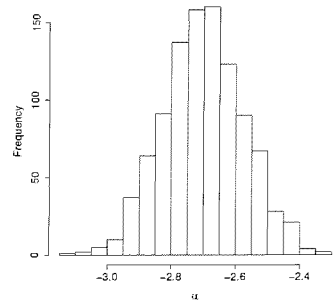
if the percentiles of the t-distribution are examined it will be noted that even for seven degrees of freedom ($n-2$ for nine GTs), the critical value is still relatively large and this does not approach an asymptote until the degrees of freedom exceed 100 (tables are given in statistical text books, such as Neter et al. (1996)). Furthermore, the ELM is typically implemented with a smaller number of targets, so that situation needs to be considered when conducting an uncertainty analysis.

Alternative approaches are adopted in sections 5.3 and 5.4, which is to implement the regression on a per-pixel basis. It is argued that these approaches should be preferred to the typical approach since they utilise all the data which is collected in the field. There also further issues that merit consideration:

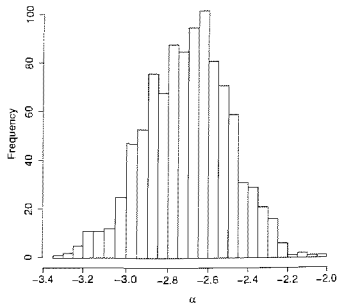
1. To implement the typical approach an average is taken over all the measurements within the target. Hence the variation in the reflectance and radiance values within the targets is lost. Strictly this means that the regression cannot then be applied to pixels that are smaller than the size of the target since a regression that is parameterised at one support may not be valid at a different support. In particular, the estimate of σ^2 is likely to decrease for models parameterised on larger supports. The magnitude of σ^2 affects the width of the confidence and prediction intervals. Generally, the user will want to conduct their atmospheric correction on pixel–sized supports.

2. The data summaries for the field-measurements of reflectance, given in section 4.5.1, raise concerns over the use of the concrete target. This histogram for band 1 shows a bimodal distribution with the remaining bands being skewed. The spatial summary plots further suggest that the concrete target can be split into two distinct sub-targets. The discussion given in section 3.2.3.2 outlines the requirement for targets to be spatially homogeneous and to be of uniform composition. The histograms and spatial summary plots show that this is not the case. This feature of the concrete target needs to be dealt with. There is a lack of accessible bright targets in this image and in the UK more generally. Hence, the user is faced with the choice of using the concrete target or being unable to implement the ELM. The methods that are presented in subsequent sections address this issue.

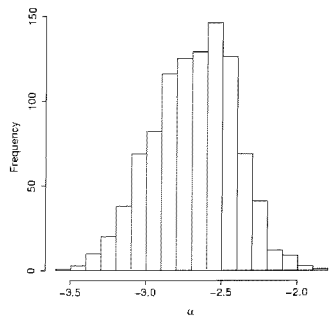
The second point is partially addressed in the next section by implementing a *modified-typical* approach. Subsequent sections consider how the data set can be used differently to provide a fuller assessment of uncertainty in the ELM and to increase the accuracy and certainty with which predictions can be made.



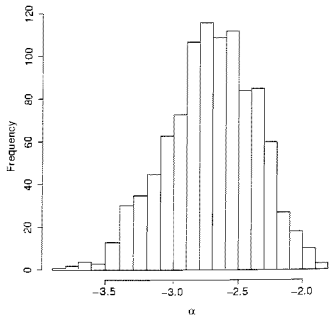
(a) $n100$



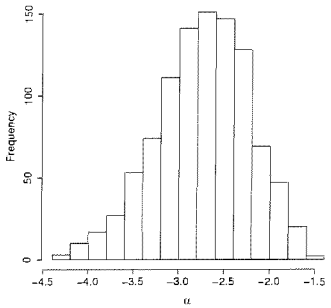
(b) $n50$



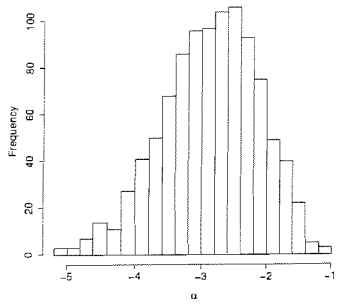
(c) $n30$



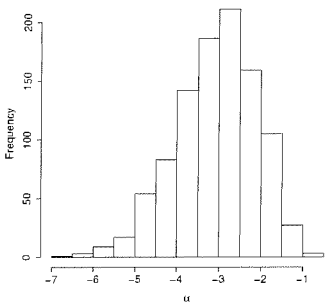
(d) $n20$



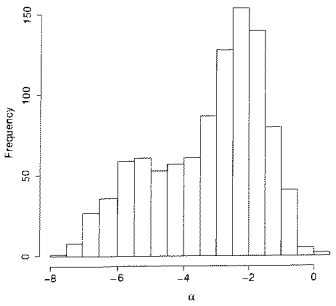
(e) $n10$



(f) $n5$

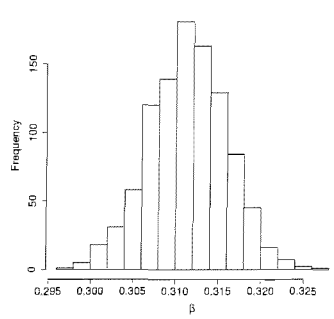


(g) $n3$

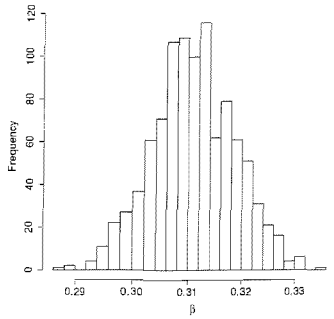


(h) $n1$

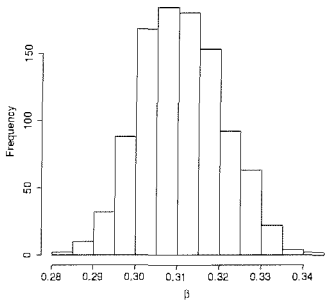
FIGURE 5.2: Histograms showing the range of estimated values for $\hat{\beta}_0$ from 1000 simulations with different sample sizes ($n100$ to $n1$).



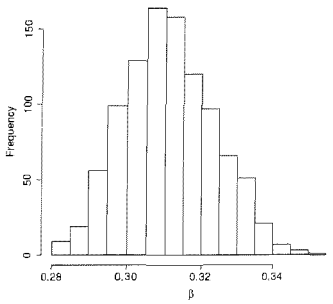
(a) $n100$



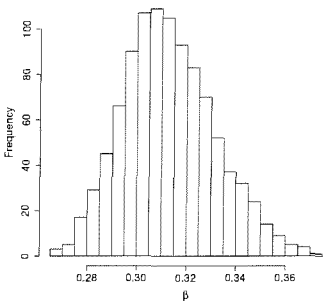
(b) $n50$



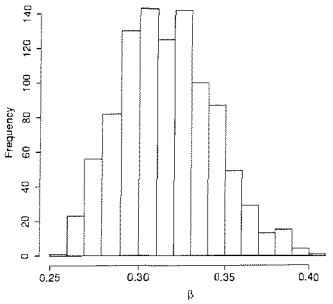
(c) $n30$



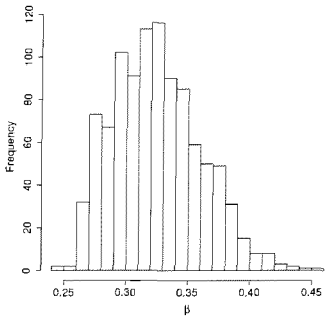
(d) $n20$



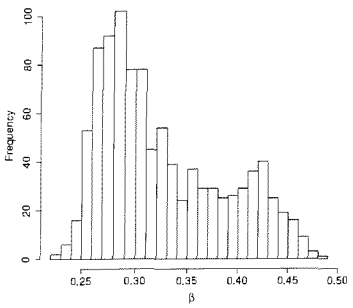
(e) $n10$



(f) $n5$

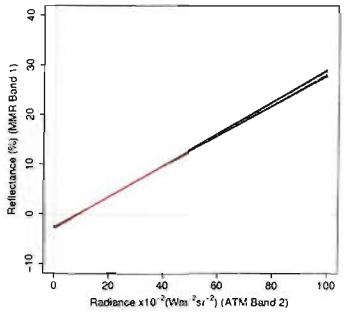


(g) $n3$

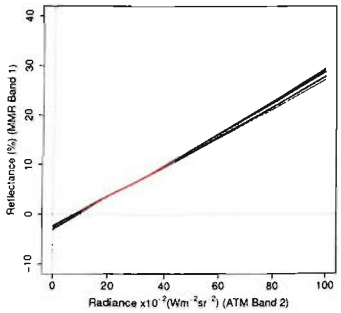


(h) $n1$

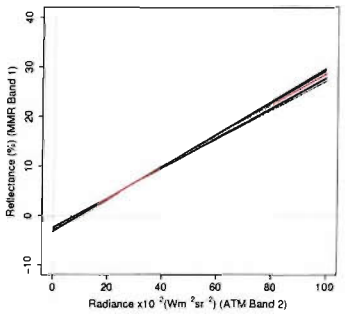
FIGURE 5.3: Histograms showing the range of estimated values for $\hat{\beta}_1$ from 1000 simulations with different sample sizes ($n100$ to $n1$).



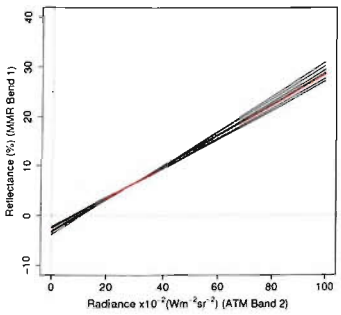
(a) $n100$



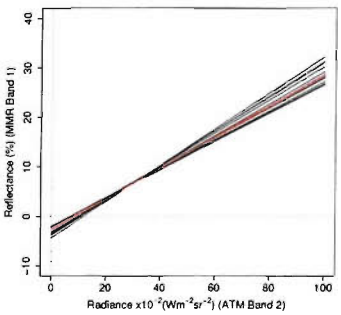
(b) $n50$



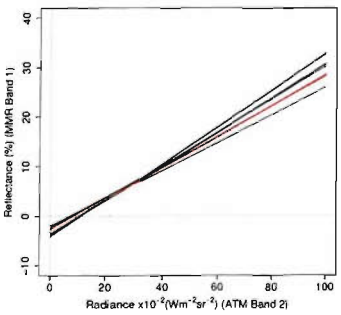
(c) $n30$



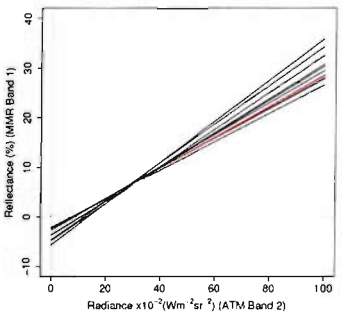
(d) $n20$



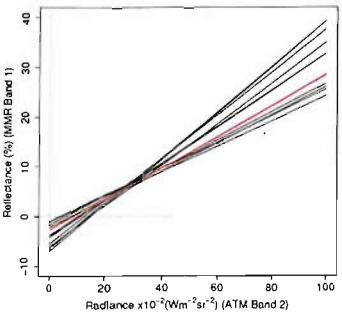
(e) $n10$



(f) $n5$

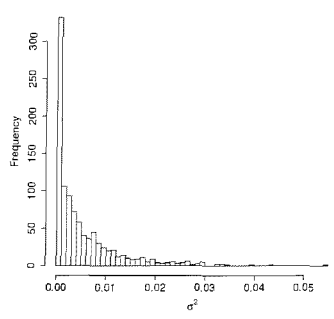


(g) $n3$

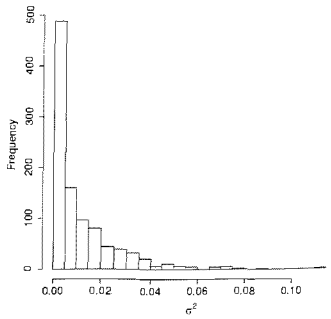


(h) $n1$

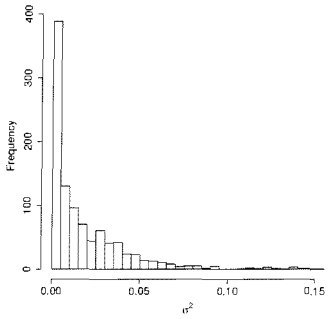
FIGURE 5.4: Plots illustrating the possible variation in the regression line for sub–samples $n100$ to $n1$. The red line shows the regression line gained for the full data set. The red line is the regression line calculated with the full data set and the black lines show the regression lines calculated from ten different realisations.



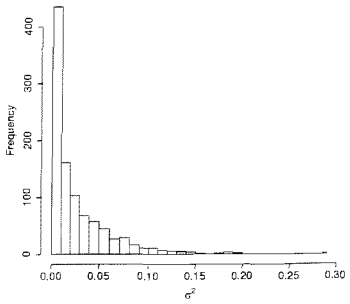
(a) $n100$



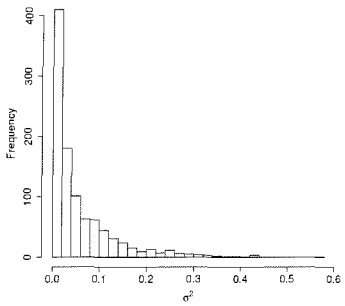
(b) $n50$



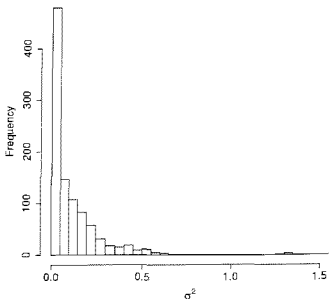
(c) $n30$



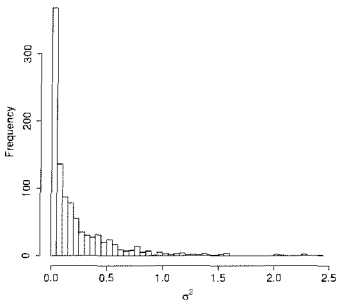
(d) $n20$



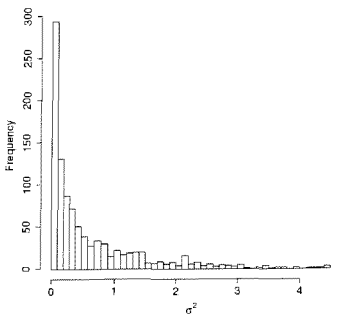
(e) $n10$



(f) $n5$



(g) $n3$



(h) $n1$

FIGURE 5.5: Histograms showing the range of estimated values for $\hat{\sigma}^2$ from 1000 simulations with different sample sizes ($n100$ to $n1$).

5.2 The modified-typical approach

The discussion at the end of the previous section highlighted the problems with the concrete target. As a result, it might be argued that the analysis given above does not reflect the usual situation for implementation of the typical approach. The analysis in this section addresses that issue by splitting the concrete target into two geographically distinct segments which better meet the related requirements of spatial homogeneity and uniform composition This better represents the typical approach that would be implemented with more appropriate targets. It is termed the *modified-typical* approach because it requires that the location of the field measurements is approximately known. Under the typical approach, the location of individual field measurements is not recorded. It is emphasised that the only reason why it is possible to implement this approach because the location of the field measurements were recorded.

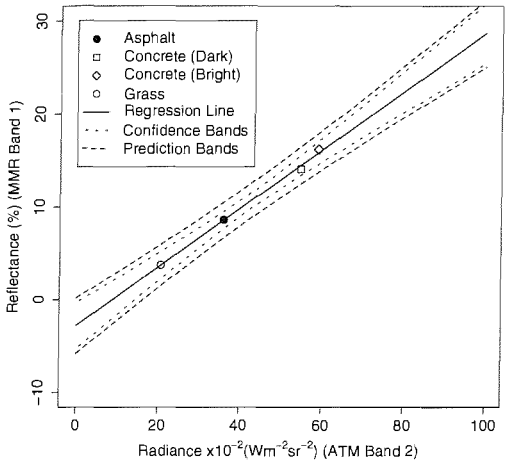
Examination of figure 4.5 showed that the north–east quadrant of the concrete target is distinctly brighter than the rest of the target. Therefore, the concrete target was segmented on this basis to produce a "bright" and "dark" concrete target. Detailed examination of the spatial structure of the GTs is given in section 5.4. The ELM was then implemented using four GTs rather than three.

The results of implementing the ELM for the four targets are given in table 5.5 and illustrated in figure 5.6. If everything else remains the same, increasing the number of data pairs will reduce the width of the confidence intervals for the parameters and for the regression line. It will also decrease the critical values for the t–test used to test whether the estimated parameters are significantly different from zero. However, unless the additional datum lies on the regression line, it will also increase $\hat{\sigma}^2$ and this may offset the effect of increasing the number of data pairs. For all four bands the t–tests showed that $\hat{\beta}_1$ was significantly different to zero at $p = 0.01$ although this was not the case for $\hat{\beta}_0$. Figure 5.6 showed that the 95% confidence intervals around the regression line were still wide.

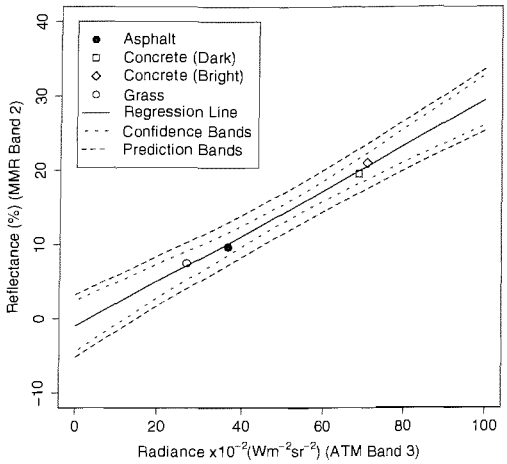
Use of the modified-typical approach led to a decrease in the width of the confidence and prediction intervals for bands 2 and 4, but an increase for bands 1 and 3. For all four bands, the width of the confidence interval around the regression line was substantial. Hence, although measures have been taken to address the problem with the concrete target, the issues are the same as they were for the typical approach, although the emergent effect are less extreme. Furthemore, this approach did not address the issue that the support of the ground targets are not all the same and that they are different to the (pixel–sized) support that is required for prediction.

	MMR B1 on ATM B2	MMR B2 on ATM B3	MMR B3 on ATM B5	MMR B4 on ATM B7
$\hat{\beta}_0$	-2.8247	-0.9687	-2.1162	-4.0784
$s.e.(\hat{\beta}_0)$	0.5758	0.7954	0.6399	1.5584
p	0.0394	0.3475	0.0805	0.1202
$p < 0.05$	Yes	No	No	No
$p < 0.01$	No	No	No	No
$\hat{\beta}_1$	0.3150	0.3042	0.3611	0.5433
$s.e.(\hat{\beta}_1)$	0.0127	0.0147	0.012	0.023
p	0.0016	0.0023	0.0011	0.0025
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\sigma}^2$	0.1527	0.3243	0.2353	0.5871
R^2	> 0.99	> 0.99	> 0.99	> 0.99

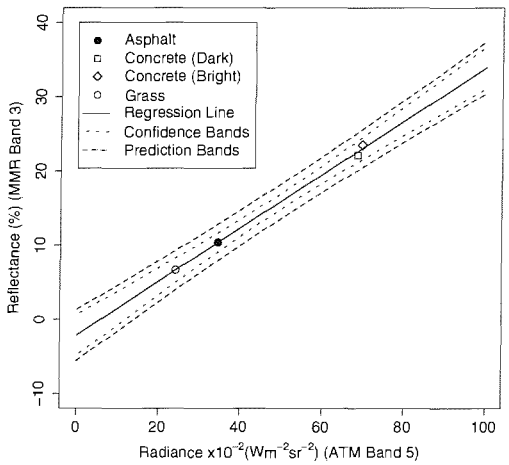
TABLE 5.5: Table showing the result of parameter estimation using of the ELM regression model using the modified-typical approach. The parameters are estimated using classical least squares and implemented in S-Plus.



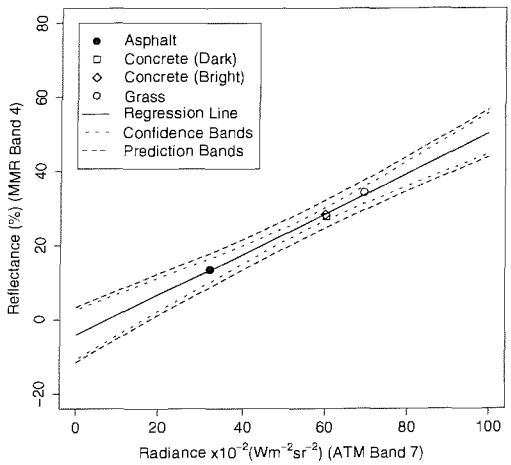
(a) ELM for MMR Band 1



(b) ELM for MMR Band 2



(c) ELM for MMR Band 3



(d) ELM for MMR Band 4

FIGURE 5.6: ELM regressions for at-surface reflectance (MMR) against at-sensor radiance (ATM) for the modified–typical approach. For each graph the regression line and 95% confidence and prediction bands are shown.

5.3 The point-pixel approach

In this section, an alternative form of the ELM is proposed and investigated. Under this approach field measurements are paired *directly* with the spatially coincident pixel. It should be noted that this is *only* possible because the location of the field measurements have been recorded. This is not something that is found in the literature. The objective is to address the following weaknesses in the typical approach:

1. Averaging the reflectance and radiance measurements for each target effectively reduces the information content by eliminating the natural variation within each target. However, if the objective is to quantify the accuracy in estimation and prediction then averaging over the targets will give a misleading result for the accuracy of the model. Furthermore, averaging over the GTs means that the number of data points for input into the regression model equals the number of GTs. Hence the degrees of freedom in the regression will be small and the t-values will be large. This is likely to lead to low confidence in the regression model.

2. When conducting regression with three or four data points then the mathematics involved mean that, if the points do not lie very close to a straight line then the confidence intervals and prediction intervals widen considerably. This leads to high uncertainty regarding the position of the regression line. Furthermore, it may also mean that hypothesis tests determine erroneously that β_1 is not significantly different to zero.

3. Various requirements for the properties of the GTs were made in section 3.2.3.2. However, finding large, spatially homogeneous GTs of uniform composition is not always feasible. These requirements are less restrictive for the point–pixel approach since the reflectance and radiance measurements are paired directly. Hence the requirements are made at the scale of the pixel rather than of the GT. For the normal error regression model, all that is required is that the $y_i|x_i$ are normally distributed.

The first two factors interact with each other. Removing the variation led to a small value for σ^2 and a large value for R^2. However, the fact that there were only three data points led to wide confidence intervals and prediction intervals. This is illustrated in the analysis presented below.

A further problem with the typical approach is that it ignores the support issue. Under the typical approach the support is not precisely specified, but it is considerably larger than the pixel size. In the typical approach implemented in section 5.1 the radiance and reflectance values were averaged over an area of approximately 35 m × 35 m, whereas the pixel size is only 2 m. Theoretically, a model that is parameterised on one support may not be applicable on another support (Blalock 1964, Heuvelink & Pebesma 1999, Bierkens

et al. 2000). However, the practical implication of this for the ELM is unclear. Adopting the point–pixel approach does not *solve* the support problem – since the reflectance and radiance data are now defined on different supports. The radiance data are defined on pixel sized supports, which are 2 m × 2 m in size. The support of the reflectance measurements is defined by the angular FOV of the instrument and the height from which the measurement was taken, and is approximately 0.3 m in diameter. However, the supports are closer in magnitude than with the typical approach so it is argued that the model is more realistic. A further assumption is that this approach assumes that the 0.3 m point is an unbiased sample of the 2 m pixel area. The support issue is addressed explicitly in section 5.4. It should be emphasised that this approach requires that the locations of the field measurements are recorded accurately.

The results for the full data set are presented in the same format as they were for the typical approach (section 5.1). Table 5.6 shows the estimated parameter values. Plots showing the fitted regression line and 95% confidence and prediction intervals are shown in figure 5.7. By comparison to the typical approach, the estimated values of β_0 and β_1 are similar. The estimated values for β_0 differ by no more than 0.5% reflectance, and the estimated values of β_1 differ by no more than 0.01. However, there are important differences between the two sets of results, as follows:

1. In all cases, the t–tests showed that the estimated values of β_0 and β_1 were significantly different to zero at $p = 0.01$. This is in contrast to the results found for the typical approach.

2. The confidence intervals on the parameter estimates and on the regression line were much narrower than for the typical and modified-typical approaches. This was clearly illustrated in figure 5.7, where the 95% confidence bands lie very close to the regression line.

3. The estimated value of σ^2 (i.e., the residual variance from the regression model) is larger for the point–pixel approach than for the typical and modified-typical approaches. This is to be expected, since the variation was not being lost in the averaging process.

4. The R^2 values were lower for the point–pixel approach than for the typical approach. For the point–pixel approach, they ranged between 0.955 and 0.974, whereas for the typical and modified-typical approaches they were all greater than 0.99. This was also due to the fact that variation was not being lost through the averaging process.

5. The confidence intervals (shown in figure 5.7) were much narrower than those found for the typical or modified-typical approaches. This may be attributed to two factors. First, because of the large number of data points, the degrees of freedom are no longer an issue, as they were with the typical approach. Second, despite the larger value of $\hat{\sigma}^2$, the data clearly support the trend indicated by the regression line. Hence the confidence intervals around the regression line are narrow.

	MMR B1 on ATM B2	MMR B2 on ATM B3	MMR B3 on ATM B5	MMR B4 on ATM B7
$\hat{\beta}_0$	-2.4164	-0.5644	-1.7760	-3.8002
$s.e.(\hat{\beta}_0)$	0.0880	0.1037	0.1045	0.2241
p	< 0.0001	< 0.0001	< 0.0001	< 0.0001
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\beta}_1$	0.3008	0.2931	0.3515	0.5410
$s.e.(\hat{\beta}_1)$	0.0021	0.0021	0.0022	0.0039
p	< 0.0001	< 0.0001	< 0.0001	< 0.0001
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\sigma}^2$	0.7814	1.1545	1.3363	2.5693
R^2	0.9663	0.9647	0.9737	0.9652

TABLE 5.6: Table showing the result of parameter estimation using of the ELM regression model using the point–pixel approach. The parameters are estimated using classical least squares and implemented in S-Plus.

6. The prediction intervals (shown in figure 5.7) were quite different to those found for the typical or modified-typical approaches. For MMR bands 1 and 3 the prediction intervals are wider than they were in the typical approach. The opposite was found for MMR bands 2 and 4. In all cases, the prediction intervals were narrower than they were for the modified-typical approach. As with the confidence intervals, the large number of data points means that the degrees of freedom are no longer an issue. Hence the prediction intervals are primarily affected by σ^2.

7. Despite the merits of this point–pixel approach it should be noted that various potential problems have been introduced.

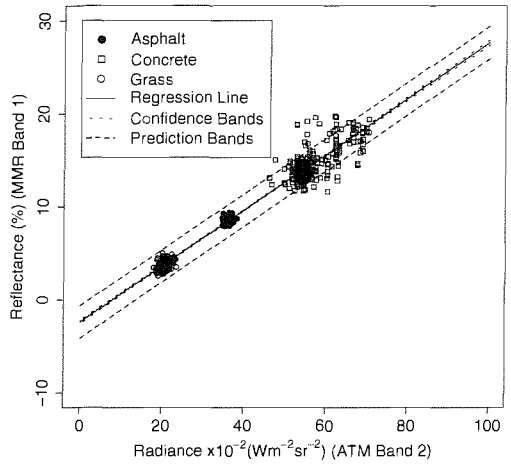
   (a) As previously stated, the point-pixel approach makes no effort to match the supports of the reflectance and radiance data. This is addressed in section 5.4.

   (b) Implementation of this approach is dependent on accurately determining the location of each field measurement. In this section, it is assumed that the determination of location is error free. However, in reality this may not be the case. This issue is addressed in chapter 6.

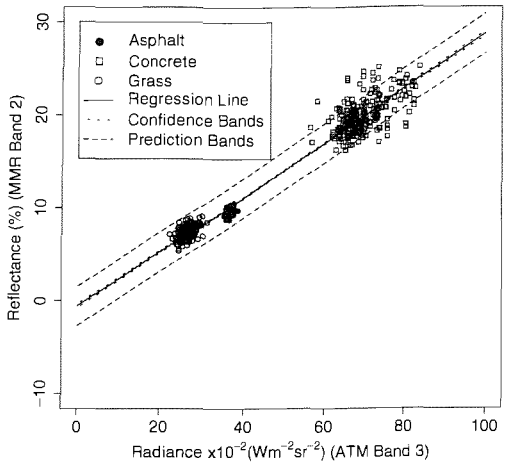
   (c) The residuals are clearly heteroskedastic. This is addressed in chapter 7.

   (d) Within each target there is spatial structure. Hence, within each target it is likely that the residuals from the regression will be autocorrelated. This is addressed in chapter 7.

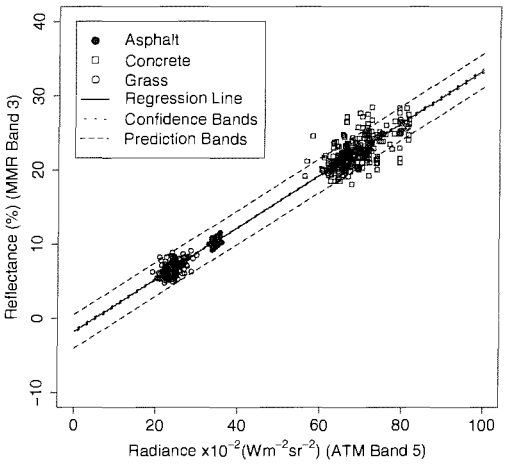
Whilst it is important to address these problems, they represent subsequent stages of the model–building process and are not addressed in this section.



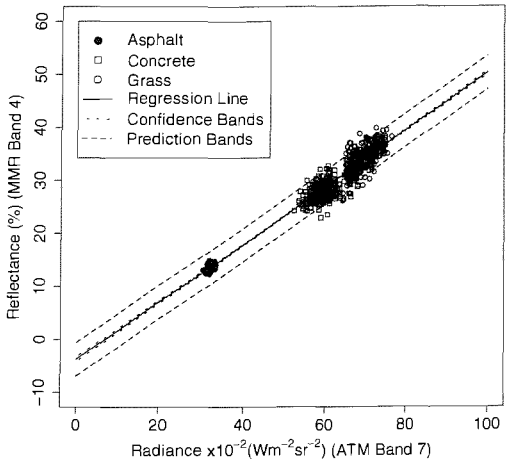
(a) ELM for MMR Band 1



(b) ELM for MMR Band 2



(c) ELM for MMR Band 3



(d) ELM for MMR Band 4

FIGURE 5.7: ELM regressions for at-surface reflectance (MMR) against at-sensor radiance (ATM) for the point–pixel approach. For each graph the regression line and 95% prediction interval are shown.

5.3.1 The implications of sample size

The implications of reducing the sample size were investigated by adopting the same procedure as was adopted for the typical approach. The same questions were addressed, as follows:

1. To what extent does reducing the sample size increase the width of the confidence interval on the parameter estimates and on the regression line? In the extreme case, this will be exemplified when the t–test shows that the estimated parameter is not significantly different to zero. As noted previously, this should not lead to the conclusion that there is no relationship be radiance and reflectance but can be

interpreted as evidence of large uncertainty in the location of the regression line.

2. To what extent does reducing the sample size increase the variability of the parameter estimates?

3. To what extent does reducing the sample size increase the size of $\hat{\sigma}^2$?

Table 5.7 shows the proportions of the simulations where $\hat{\beta}_0 \neq 0$ and $\hat{\beta}_1 \neq 0$ for the different sample sizes. In all cases *except* for $n1$ it was found that $\hat{\beta}_1 \neq 0$ at $p = 0.01$. However, the situation was quite different for β_0. For MMR bands 1, 3 and 4 at least 20 measurements per target were required to ensure that $> 99\%$ of realisations led to $\beta_0 \neq 0$ at $p = 0.01$. For MMR band 2, the full data set was required, although $\hat{\beta}_0 = -0.56$ for *nFull*. This is much closer to zero than the values of $\hat{\beta}_0$ for the other wavebands. *If* this magnitude of bias is acceptable to the user then they could collect 20 measurements per target in order to cover estimation of β_0 in the other wavebands, and disregard the estimation of β_0 for MMR band 2.

The above analysis was useful because it indicated the approximate sample size required to obtain a regression that was consistently statistically significant. However, considering only statistical significance can lead to misleading conclusions (Cox 2001, Sterne & Davey Smith 2001). In this case, the t–test indicated that the estimated parameters were significantly different from zero but says little about the uncertainty attached to parameter estimates. Sterne & Davey Smith (2001) emphasised the need to consider significance tests in context and to also consider the confidence limits or credible regions around the estimated parameter.

The issue of variability in the parameter estimates and regression line was investigated for MMR band 1. The range of variability in $\hat{\beta}_0$ and $\hat{\beta}_1$, gained from drawing 1000 realisations for a given sample size, are illustrated using histograms in figures 5.8 and 5.9 respectively. It is clear that the variability in the parameter estimates increased as the number of measurements per target decreased.

The effect of sample variability on estimation of the regression line is shown in figure 5.10. This shows the regression line for the full data set (red line) and the regression line calculated on the basis of ten realisations (black lines). Subjectively, the range of variability does not seem to be as large as it was for the typical approach. This is borne out by table 5.8 which shows the maximum and minimum predicted values at $x = 40$ and $x = 50$. If the user could tolerate a range of 1% (reflectance), then they would require a sample size of 20 measurements per target at $x = 40$ and 50 measurements per target at $x = 50$. It is emphasised that this discussion applies only to the estimated regression line. It is also important to consider the uncertainty around the regression line. This can be illustrated by the confidence interval around the regression line. This is shown, for one realisation, in figure 5.11 for n100, n50, n30 and n20. These figures show that the confidence interval on the regression tended to increase as the sample size decreased and it was visually substantially narrower for $n \geq 50$.

The above discussion was useful for emphasising that variability in the regression increases as sample size decreases. However, to gain a fuller picture it is also necessary to consider the variability of σ^2. This was illustrated in figure 5.12. It should be recognised that $\hat{\sigma}^2 \sim \chi^2_{n-2}$ (where n is the sample size) for all sample sizes (Casella & Berger 1990, Neter et al. 1996). For large sample sizes (e.g., $nFull$ and $n100$) this will approximate a normal distribution. However, for smaller sample sizes, the distribution is clearly non–normal. For smaller sample sizes, the sample mean value of $\hat{\sigma}^2$ approximated that obtained for $n100$ or the full data set. However, the shape of the distribution means that there is a high probability of underestimating σ^2. This means that the confidence intervals and prediction intervals may actually be underestimated for smaller samples. Furthermore, as the sample size decreases, the probability of estimating extremely large values increases.

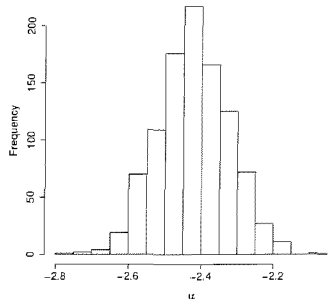
The above discussion on the variability of $\hat{\sigma}^2$ suggests caution when interpreting the results from figures 5.8, 5.9 and table 5.8. In particular, it suggests that, to maintain a relatively low range in the estimated values of σ^2 a relatively large sample size is required. Hence, the recommendation is that at least 50 (and preferably 100) measurements should be obtained per data set.

5.3.2 The point-pixel approach: conclusions and recommendations

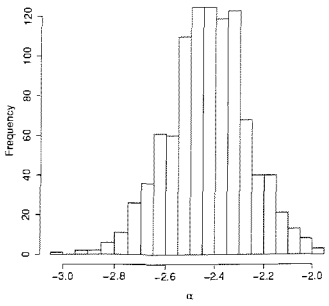
The above discussion showed that 20 measurements per target were required in order to ensure that β_0 and β_1 were significantly different to zero at $p = 0.01$ for all wavebands *except* MMR band 2. Furthermore, it was demonstrated that the information contained in the simulations could be used to specify a sample size to satisfy a given tolerance for the regression line. However, the necessity to consider the confidence intervals around the parameter estimates and regression line was also emphasised and illustrated in figure 5.11. However, in addressing this, the stability of the estimate of σ^2 also needs to considered. The variability in the estimate of σ^2 increases substantially where less than 50 measurements per target are taken.

The above results suggested that 20 measurements per target (i.e., a total of sixty measurements) is the minimum that is viable. However, given the importance of accurate atmospheric correction for remote sensing, more measurements will be required to ensure that $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ are not sensitive to sample variability and to ensure narrow confidence bands around the regression. Hence, a minimum of 50 measurements per target is recommended and at least 100 would be preferable.

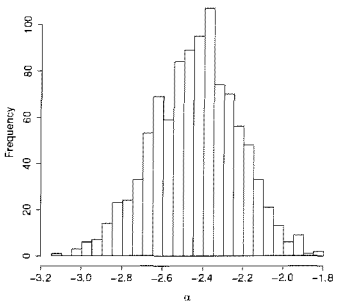
In addition to the support issue there are two problems with the regression model that has been implemented here. First, the residuals are clearly heteroskedastic (see figure 5.7). Second, the y_i and x_i are not independent within each target. These two issues are addressed in chapter 7. The next section will address the support issue by aggregating the reflectance data to pixel size supports.



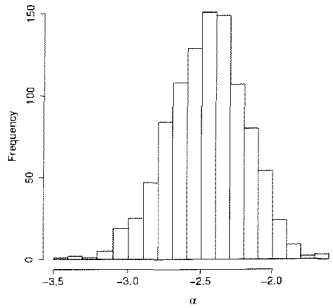
(a) $n100$



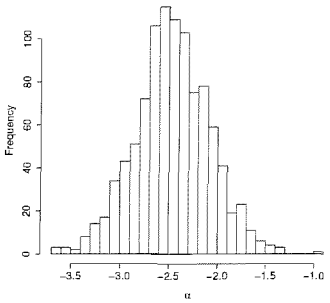
(b) $n50$



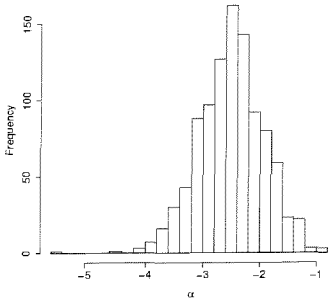
(c) $n30$



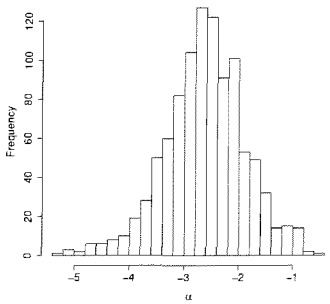
(d) $n20$



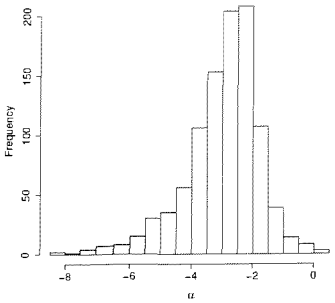
(e) $n10$



(f) $n5$

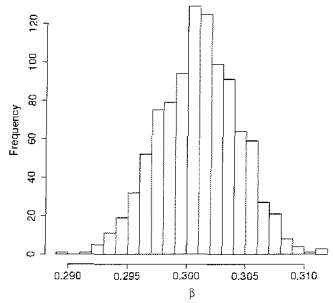


(g) $n3$

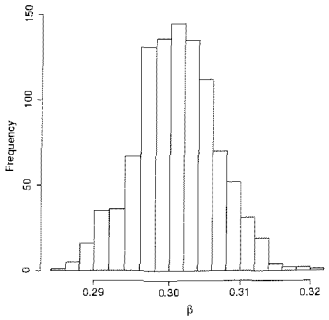


(h) $n1$

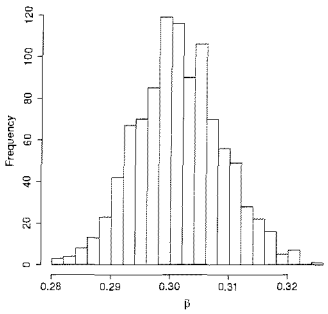
FIGURE 5.8: Histograms showing the range of estimated values for $\hat{\beta}_0$ from 1000 simulations with different sample sizes ($n100$ to $n1$, point-pixel approach).



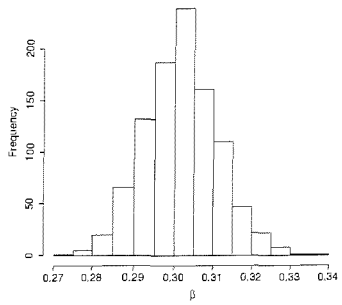
(a) $n100$



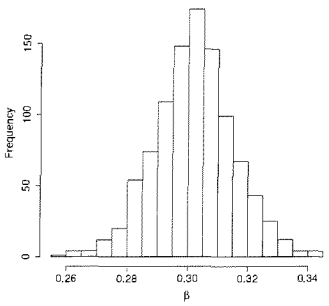
(b) $n50$



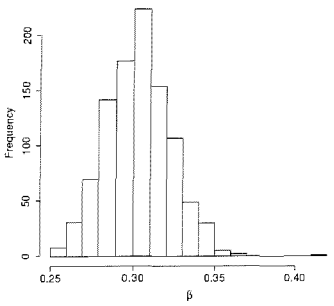
(c) $n30$



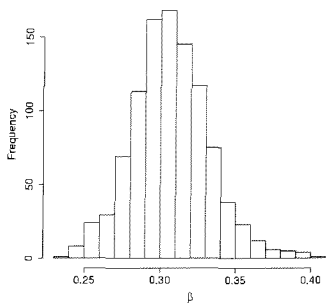
(d) $n20$



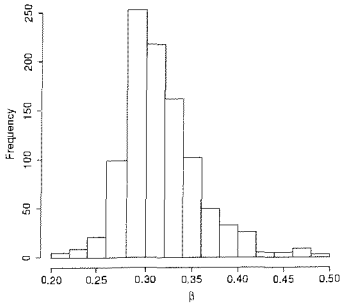
(e) $n10$



(f) $n5$

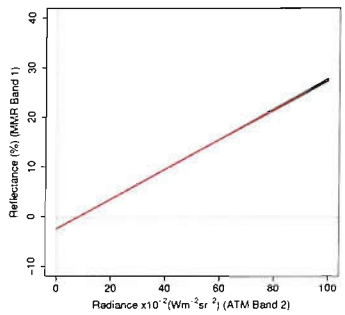


(g) $n3$

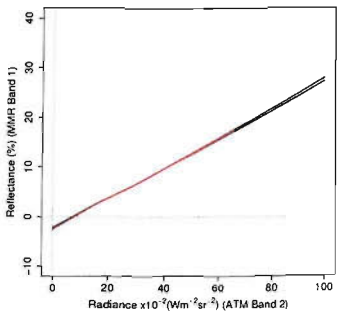


(h) $n1$

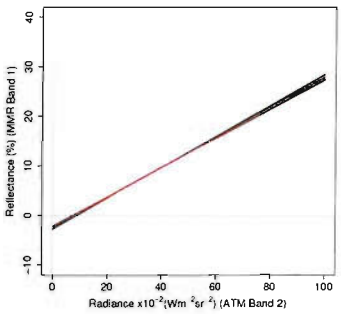
FIGURE 5.9: Histograms showing the range of estimated values for $\hat{\beta}_1$ from 1000 simulations with different sample sizes ($n100$ to $n1$, point-pixel approach).



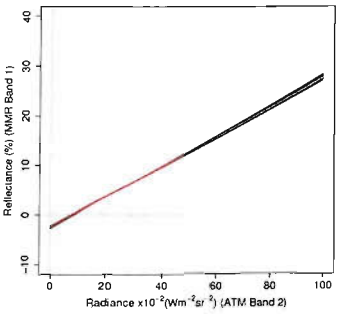
(a) $n100$



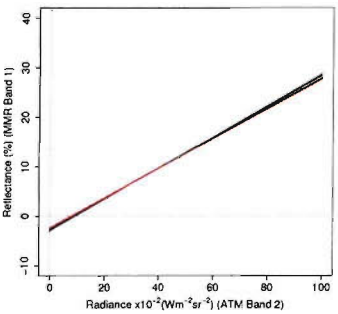
(b) $n50$



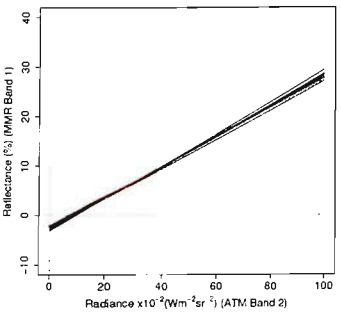
(c) $n30$



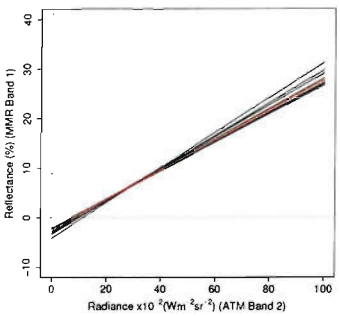
(d) $n20$



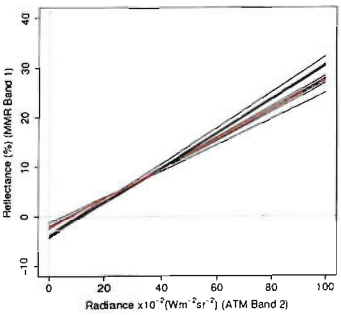
(e) $n10$



(f) $n5$

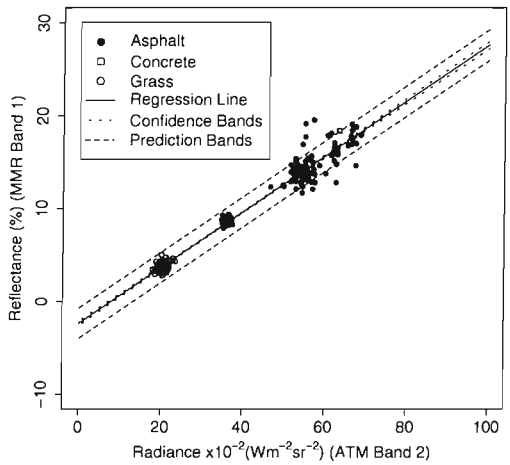


(g) $n3$

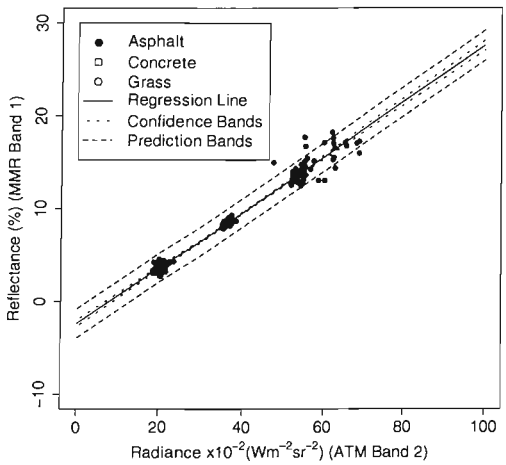


(h) $n1$

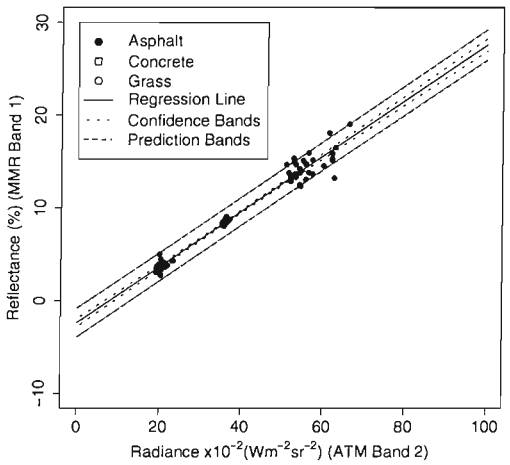
FIGURE 5.10: Graphs illustrating the possible variation in the regression line for sub-samples $n100$ to $n1$ (point–pixel approach). The red line is the regression line calculated with the full data set and the black lines show the regression lines calculated from ten different realisations for each given sample size.



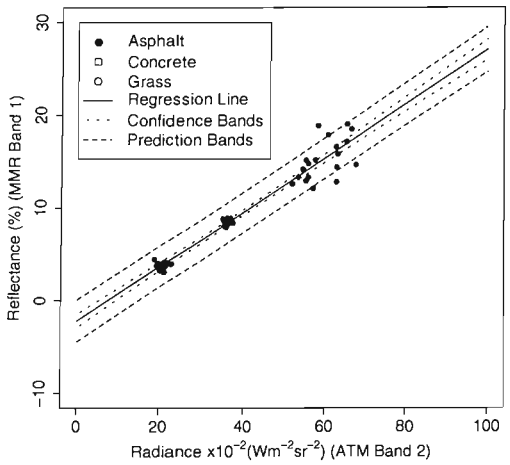
(a) 100 measurements per target.



(b) 50 measurements per target.



(c) 30 measurements per target.



(d) 20 measurements per target.

FIGURE 5.11: ELM regressions for at-surface reflectance (MMR band 1) against at-sensor radiance (ATM band 2) calculated using the point-pixel approach for different sample sizes. For each graph the regression line and 95% prediction interval are shown.

(a) Parameter estimation for y = MMR band 1 and x = ATM band 2.

	$p < 0.05$		$p < 0.01$	
Sample size	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$
$nFull$	100	100	100	100
$n100$	100	100	100	100
$n50$	100	100	100	100
$n30$	100	100	100	100
$n20$	100	100	100	100
$n10$	100	100	99.2	100
$n5$	96.2	100	82.9	100
$n3$	86.3	100	65.9	100
$n1$	21.7	79.5	4.0	20.8

(b) Parameter estimation for y = MMR band 2 and x = ATM band 3.

	$p < 0.05$		$p < 0.01$	
Sample size	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$
$nFull$	100	100	100	100
$n100$	99.9	100	98.6	100
$n50$	88.5	100	67.3	100
$n30$	65.1	100	37.1	100
$n20$	50.4	100	22.2	100
$n10$	25.9	100	8.5	100
$n5$	15.5	100	3.9	100
$n3$	13.9	100	2.6	100
$n1$	3.3	55.3	0.5	7.9

TABLE 5.7: Tables showing the proportion of the parameter estimates that are significantly different to zero for each sample size (1000 realisations per sample size, point–pixel approach). The t-tests are evaluated at the 95% (p=0.05) and 99% (p=0.01) confidence levels.

(c) Parameter estimation for y = MMR band 3 and x = ATM band 5.

Sample size	$p < 0.05$		$p < 0.01$	
	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$
$nFull$	100	100	100	100
$n100$	100	100	100	100
$n50$	100	100	100	100
$n30$	100	100	100	100
$n20$	100	100	99.5	100
$n10$	95.3	100	83.2	100
$n5$	74.0	100	50.3	100
$n3$	59.3	100	33.1	100
$n1$	15.2	88.0	2.6	23.5

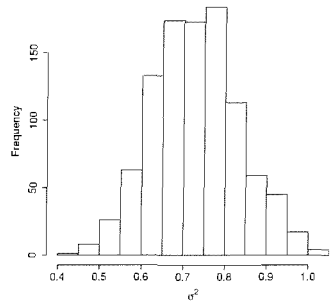
(d) Parameter estimation for y = MMR band 4 and x = ATM band 7.

Sample size	$p < 0.05$		$p < 0.01$	
	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$	% of samples where $\beta_0 \neq 0$	% of samples where $\beta_1 \neq 0$
$nFull$	100	100	100	100
$n100$	100	100	100	100
$n50$	100	100	100	100
$n30$	100	100	100	100
$n20$	100	100	100	100
$n10$	99.8	100	96.6	100
$n5$	83.9	100	47.5	100
$n3$	52.4	100	17.6	100
$n1$	8.5	52.1	1.9	12.5

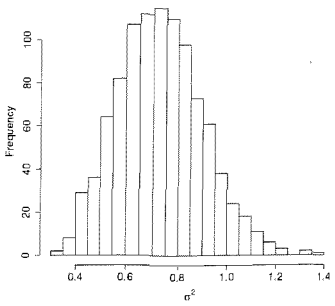
TABLE 5.7: continued

Sample Size	Predicted Value ($X = 40$)		Predicted Value ($X = 50$)	
	Minimum	Maximum	Minimum	Maximum
$n100$	9.4	9.8	12.4	12.9
$n50$	9.4	9.9	12.3	13.1
$n30$	9.3	10.0	12.1	13.2
$n20$	9.2	10.1	12.0	13.5
$n10$	9.0	10.1	11.8	13.5
$n5$	8.9	10.9	11.4	15.1
$n3$	8.6	10.8	11.1	14.8
$n1$	8.3	12.0	10.4	16.8

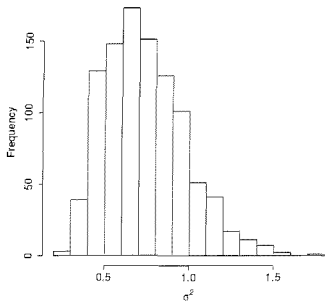
TABLE 5.8: Table showing the range of predicted regression values for $x = 40$ and $x = 50 \times 10^{-2}\mathrm{Wm}^{-2}\mathrm{sr}^{-2}$ for MMR band 1 regressed on ATM band 2 (point-pixel approach) and different sample sizes.



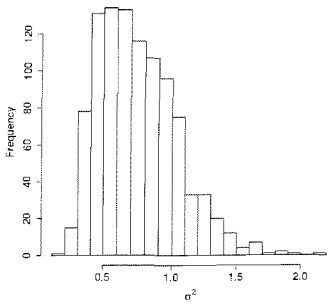
(a) $n100$



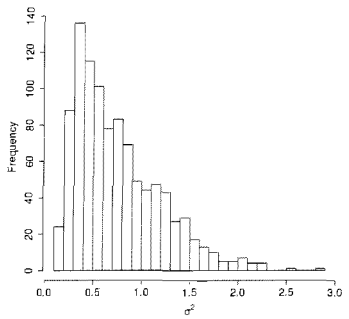
(b) $n50$



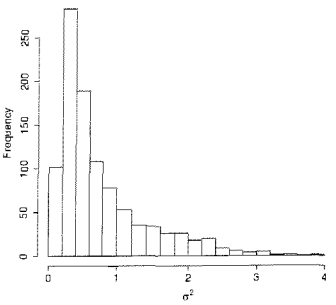
(c) $n30$



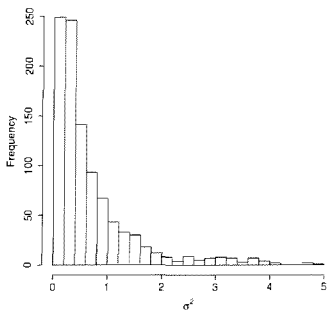
(d) $n20$



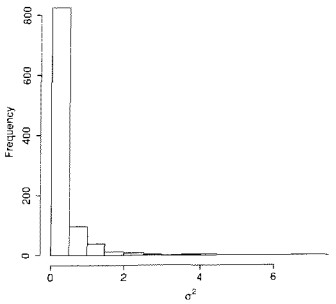
(e) $n10$



(f) $n5$



(g) $n3$



(h) $n1$

FIGURE 5.12: Histograms showing the range of $\hat{\sigma}^2$ from 1000 simulations with different sample sizes ($n100$ to $n1$, point-pixel approach).

5.4 The block-pixel approach

In this section, the approaches used to aggregate the field data to the same support as the remotely sensed data are discussed and the analysis is presented. A more general discussion of scale and the spatial support, together with some background on aggregation is given in section 2.2.1. An approach based on simple averaging is first considered (section 5.4.1), followed by geostatistical approaches (section 5.4.2).

5.4.1 Simple averaging

For the analysis in this section, a simple approach to aggregation was adopted. This was achieved by averaging all the field measurements that were made within any one pixel. This is analogous to the design–based methods described by Bierkens et al. (2000). A potential criticism of this approach is that design-based methods are generally based on random sampling (Brus & de Gruijter 1997, Bierkens et al. 2000), whereas a gridded sampling scheme was used. A further criticism might be that the number of field measurements per pixel was low (less than four for a 2 m × 2 m pixel). Addressing both these issues would not be feasible logistically in the ELM context. This is because a random sample design would take longer and be more cumbersome to implement. This would be exacerbated if a larger sample were taken. However, obtaining a large random sample may be feasible for other empirical models.

Despite the limitations outlined above this approach was straight forward to implement and addressed the support issue by averaging out within–pixel variation. The procedure was executed as follows:

1. Match each field measurement with its spatially coincident pixel measurement.
2. Take the average of all field measurements taken within each pixel and pair this with the remotely sensed radiance measurement for that pixel.
3. Exclude all pixels with zero or one measurement of reflectance. This leaves a total of 36 data pairs for asphalt and concrete and 39 for grass (111 data pairs in total).
4. Compute the regression using the remaining data pairs.

The results for the implementation of the regression line are shown in table 5.9 and figure 5.13. It is useful to note that the estimated value of σ^2 was lower than it was for the point-pixel approach. This was attributed to the fact that variability was lost in the averaging process. This was desirable, since reflectance was averaged on a per-pixel basis – which is consistent with desired support size for prediction. If n remains constant, a reduction in $\hat{\sigma}^2$ leads to a reduction in the width of the confidence and prediction intervals. This effect is largest when the degrees of freedom are very small (10 or less) and reduces

for larger degrees of freedom. For $n = 111$, the width of the confidence and prediction intervals was controlled by the size of σ^2 rather than n. Hence, the reduction in $\hat{\sigma}^2$ relative to the point-pixel approach led to a reduction in the uncertainty in parameter estimation and new variable prediction.

When this approach is considered relative to the typical approach it will be noted that σ^2 is larger, but the confidence intervals around the regression line are narrower.

It might be argued that step 3 in the above procedure could be omitted. The assumption (as with the point-pixel approach) is that an individual measurement represents an unbiased sample of the entire pixel. The results for this approach are shown in table 5.10. This approach yields a total of 436 data pairs to support the regression model. However, of these 436 pixels, 325 contain *one* field measurement – so there is no averaging across multiple data points. Hence, $\hat{\sigma}^2$ was larger; however, the estimates of β_0 and β_1 are similar.

Overall, it is preferable to include *only* those pixels which contain two or more field-measurements. This explicitly addresses the support issue and hence reduces the sensitivity of the ELM to sub-pixel variation in reflectance. Furthermore, as discussed in chapter 6, it reduces the sensitivity of the ELM to positional uncertainty in the field measurements. However, in this case, because of the sample structure, it leads to substantial redundancy. A sampling strategy designed specifically for the simple averaging approach could limit this. Hence, the results from both approaches are presented here.

Adopting the simple averaging approach explicitly addressed the support issue. The advantages of this approach were, firstly, its simplicity and, second, it is not based on any assumption about the underlying spatial variability. The disadvantage, was that sampling was not conducted at random. Hence this approach cannot be considered design–based. This method was, also, limited by uncertainty in the location of the field measurements. Furthermore, if this approach was to be applied operationally it is recommended that a sampling design be constructed that would minimise the number of pixels with only one measurement in them.

5.4.2 Block kriging

In this section, the geostatistical approach of *block-kriging* is used to aggregate from the field measurements (approximated by a point) to pixel-sized blocks. A detailed discussion of the approach for MMR band 1 (paired with ATM band 2) is given. Each target is dealt with in turn and the ELM regression model is then applied.

The notation used in the geostatistical analysis is listed as follows:

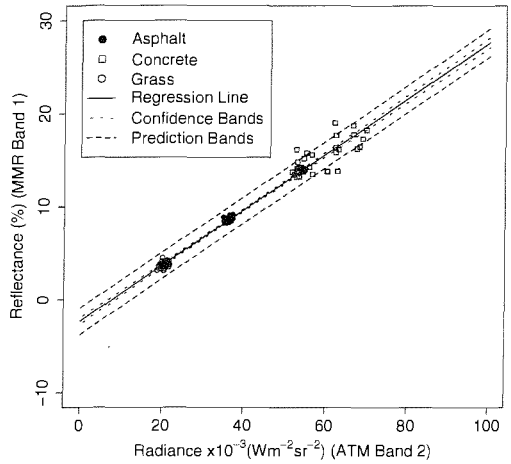
- τ_n^2 refers to the nugget

	MMR B1 on ATM B2	MMR B2 on ATM B3	MMR B3 on ATM B5	MMR B4 on ATM B7
$\hat{\beta}_0$	-2.3367	-0.4216	-1.6934	-3.8556
$s.e.(\hat{\beta}_0)$	0.1742	0.2067	0.1907	0.4238
p	< 0.00001	0.0438	< 0.00001	< 0.00001
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	No	Yes	Yes
$\hat{\beta}_1$	0.2997	0.2903	0.3505	0.5454
$s.e.(\hat{\beta}_1)$	0.0043	0.0043	0.0041	0.0075
p	< 0.00001	< 0.00001	< 0.00001	< 0.00001
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\sigma}^2$	0.4981	0.7412	0.7218	1.6496
R^2	0.9785	0.9764	0.9853	0.9799

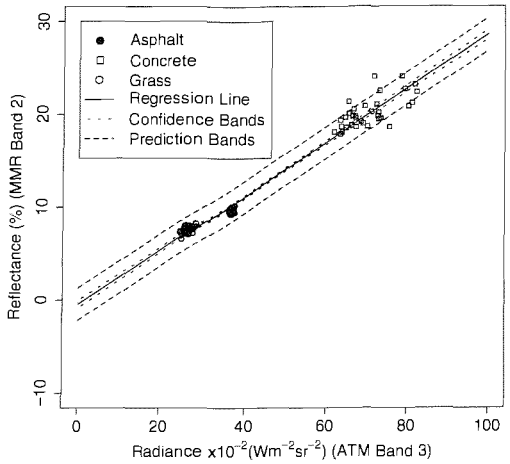
TABLE 5.9: Table showing the result of parameter estimation of the ELM regression model using the simple averaging approach. All pixels containing *two or more* measurements are included. The parameters were estimated using classical least squares and implemented in S-Plus.

	MMR B1 on ATM B2	MMR B2 on ATM B3	MMR B3 on ATM B5	MMR B4 on ATM B7
$\hat{\beta}_0$	-2.483	-0.6142	-1.8526	-3.8777
$s.e.(\hat{\beta}_0)$	0.0987	0.1208	0.1212	0.2778
p	< 0.00001	< 0.001	< 0.00001	< 0.00001
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\beta}_1$	0.3021	0.2949	0.3529	0.5415
$s.e.(\hat{\beta}_1)$	0.0024	0.0025	0.0025	0.0048
p	< 0.00001	< 0.00001	< 0.00001	< 0.00001
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\sigma}^2$	0.6121	0.9721	1.1128	2.262
$\hat{\sigma}$	0.7824	0.9859	1.055	1.504
R^2	0.9738	0.9709	0.9785	0.9671

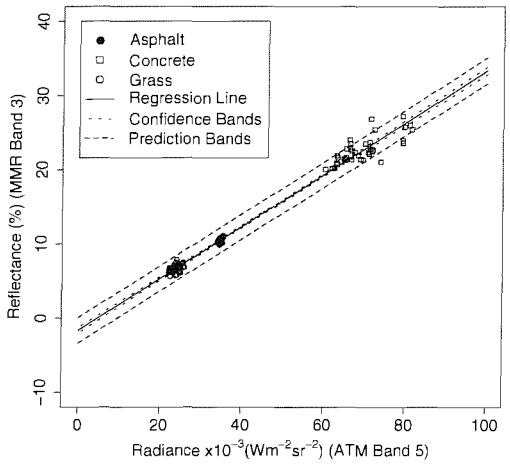
TABLE 5.10: Table showing the result of parameter estimation of the ELM regression model using the simple averaging approach. All pixels containing *one or more* field measurement are included. The parameters were estimated using classical least squares and implemented in S-Plus.



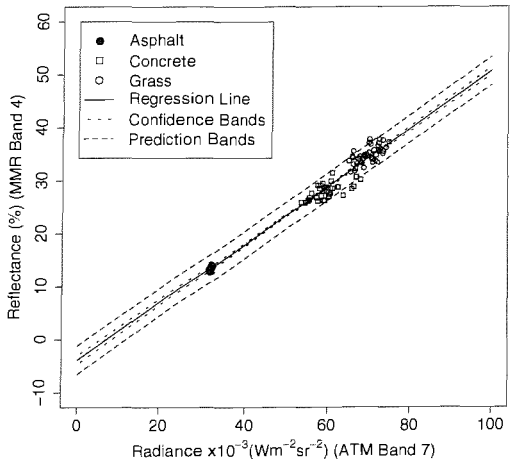
(a) ELM for MMR band 1



(b) ELM for MMR band 2



(c) ELM for MMR band 3



(d) ELM for MMR band 4

FIGURE 5.13: ELM regressions for at–surface reflectance (MMR) against at–sensor radiance (ATM) for the simple averaging approach (two or more measurements per pixel). For each graph the regression line and 95% confidence and prediction bands are shown.

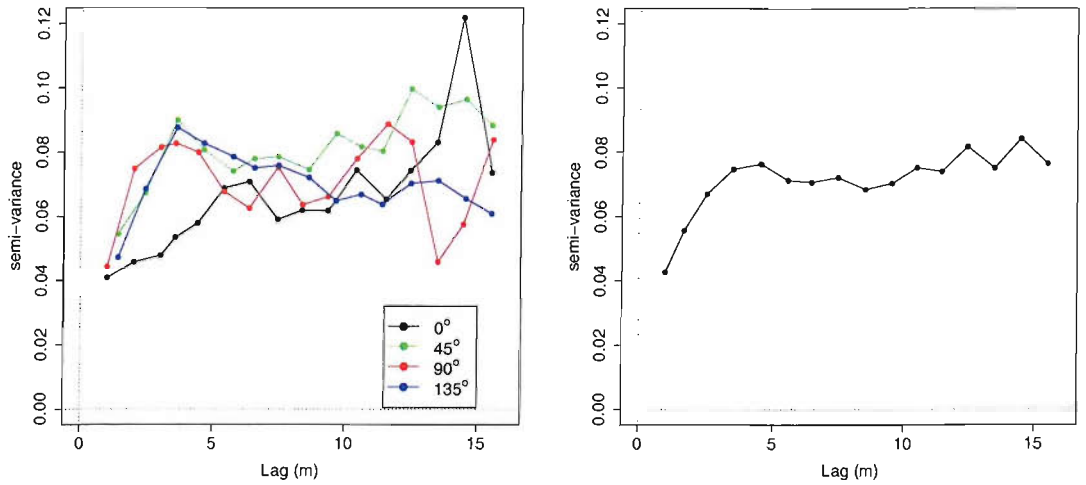
- τ_s^2 refers to the partial sill. Hence $\sigma^2 = \tau_n^2 + \tau_s^2$. *This differs from its use in the rest of the chapter, where it refers to the residual variance in the regression model.* The sill is therefore $\tau_n^2 + \tau_s^2$. From section 5.4.2.5 usage is returned to meaning the residual variance (from the regression model).

- ϕ is the variogram range.

- ME refers to the mean error. It is a measure of bias.

- MSE refers to the mean square error. It is a measure of accuracy.

- MSDR is the mean square deviation ratio. A value close to one indicates that the kriging variance is consistent with the cross-validation variance.



(a) Directional variogram

(b) Omnidirectional variogram

FIGURE 5.14: Directional and omnidirectional sample variograms for MMR band 1 over the asphalt surface.

- SS refers to to the sum of square errors.
- AIC refers to the Aikaike information criterion (AIC).

5.4.2.1 MMR band 1: asphalt GT

As discussed in section 4.5.1, the spatial summary plot given in figure 4.5(a) demonstrated the presence of spatial structure. Geostatistics can be used to model this variation and to predict a surface using point or block supports.

Under the conventional geostatistical approach it is usual to model the variogram up to half the maximum lag (32 m in this case). Given this restriction it was sensible to perform kriging using a neighbourhood that was no larger than 16 m. Directional and omnidirectional variograms are shown in figure 5.14. The directional variogram showed that the nature of the spatial variation was consistent in all but one direction. Hence although there is some suggestion of anisotropy the evidence is not strong and this was not explored further. Furthermore, the spatial summary plot (figure 4.5(a)) showed evidence for a trend, although evidence for this is not manifested in the sample variograms. This left several choices about how best to model the variogram. The variogram was, therefore, modelled with and without a trend and the resulting cross-validation statistics compared.

The first stage was to model the variogram without accounting for the trend. The shape of the variogram showed that the semi-variance increased sharply and then levelled off. Given this shape, the circular, spherical and exponential models were fitted. The exponential model was least accurate. Descriptive statistics for the circular and spherical models are shown in table 5.11. These showed that there was no strong evidence to

	Variogram				Cross–validation				
Model	τ_n^2	τ_s^2	ϕ	SS	ME	MSE	MSDR	*Slope*	R^2
Spherical	0.02	0.05	3.98	0.30	-0.006	0.05	1.02	0.29	0.29
Circular	0.0	0.05	3.62	0.30	-0.006	0.05	0.99	0.30	0.30

TABLE 5.11: Comparison of different variogram models fitted to MMR band 1 over the asphalt surface. The variogram model was fitted to the sample variogram and cross–validation prediction conducted using OK.

	Variogram				Cross–validation				
Model	τ_n^2	τ_s^2	ϕ	SS	ME	MSE	MSDR	*Slope*	R^2
Spherical	0.015	0.049	3.16	0.15	0.001	0.05	1.14	0.30	0.26
Circular	0.018	0.046	2.91	0.15	0.001	0.05	1.15	0.30	0.26

TABLE 5.12: Comparison of different variogram models fitted to MMR band 1 over the asphalt surface. The variogram model is fitted to the sample variogram after first fitting a first–order trend on location. Cross–validation prediction was conducted using KT.

choose between the circular and spherical model. Note that the slope and R^2 for the cross–validation were small. This may partly reflect the lack of data at short lags.

Table 5.12 illustrates the situation when a first order trend on the location was included. The results from the cross–validation showed that, although spatial data summaries showed evidence for a trend, including the trend did not increase the accuracy of the model. Hence, the circular model *without* the trend was adopted.

The MBG maximum likelihood approach was also considered. The results are shown in tables 5.13 and 5.14 for the fitting of models with and without a trend, respectively. These results suggested that fitting the trend increased the accuracy of the model. The AIC showed that the increased accuracy justified the decrease in parsimony. The spherical and circular model were more accurate than the exponential model. The AIC suggested that the circular model provided a more accurate fit to the data than the spherical model, although the cross-validation showed that adopting the spherical model would not greatly decrease the accuracy of prediction. The circular model was adopted. It should be noted that when the maximum likelihood approach is used the model is fitted to all the data, whereas with the conventional approach the model was not fitted at the long lags. Hence, under the MBG–ML approach it was necessary to account explicitly for the trend when fitting the model.

5.4.2.2 MMR band 1: concrete GT

An initial assessment of the spatial summary plots (section 4.5.1) suggested that the concrete target was composed of two sub targets. The concrete is actually laid in regular 6 m × 6 m slabs. Figure 5.15 shows a spatial summary plot of the concrete surface. The size of the symbol is proportional to the magnitude of the reflectance measurement.

	Variogram				Cross–validation				
Model	τ_n^2	τ_s^2	ϕ	AIC	ME	MSE	MSDR	*Slope*	R^2
Spherical	0.02	0.05	3.92	0.9	-0.003	0.05	1.03	0.27	0.25
Circular	0.02	0.05	3.66	-1.3	-0.003	0.05	1.03	0.28	0.27
Exponential	0.02	0.50	2.86	-4.07	-0.0004	0.05	0.92	0.23	0.20

TABLE 5.13: Comparison of different variogram models fitted to MMR band 1 over the asphalt surface. The model was fitted using ML without a trend and cross–validation prediction conducted using OK.

	Variogram				Cross–validation				
Model	τ_n^2	τ_s^2	ϕ	AIC	ME	MSE	MSDR	*Slope*	R^2
Spherical	0.02	0.04	3.79	-15.8	-0.0008	0.05	1.04	0.30	0.30
Circular	0.02	0.04	3.57	-17.3	-0.0006	0.05	1.01	0.30	0.30
Exponential	0.01	0.05	1.23	-13.1	0.01	0.06	0.97	0.12	0.17

TABLE 5.14: Comparison of different variogram models fitted to MMR band 1 over the asphalt surface. The model was fitted using ML and a first order trend on location. Cross–validation prediction was conducted using KT.

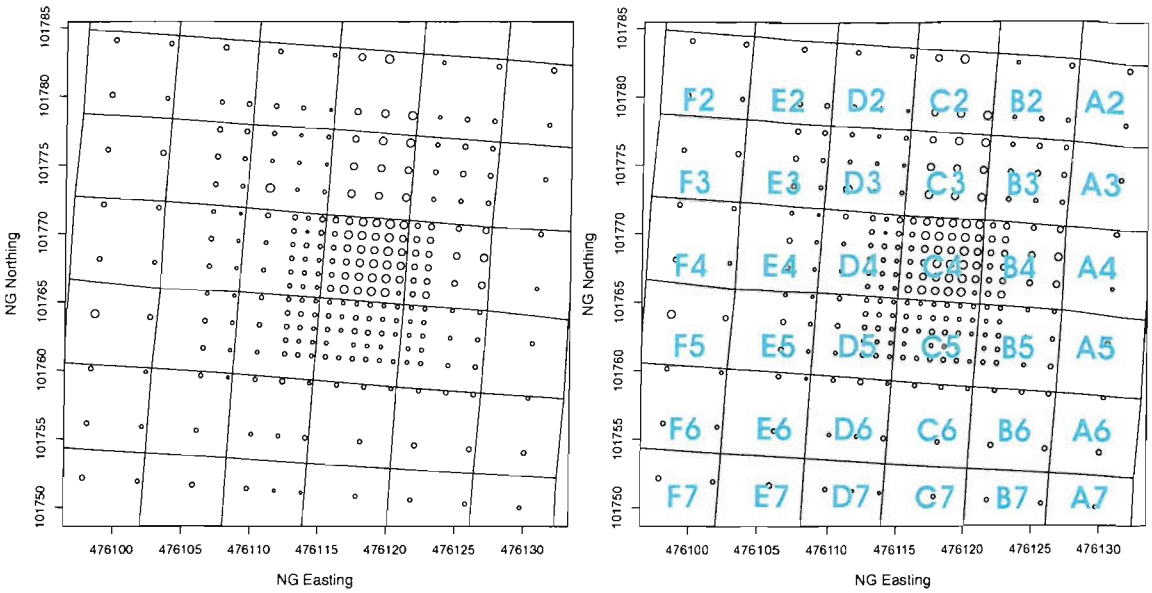
Figure 5.15(b) shows the slabs, which are labelled according to the author's co–ordinate system. This suggests that slabs A2, A3, A4, B2, B3, B4, C2, C3 and C4 represent a clearly separable surface. The measurements taken over these slabs were segmented and labelled "BC" (bright concrete, 87 measurements), the remaining slabs were labelled "DC" (dark concrete, 175 measurements). When the concrete GT was segmented on this basis, the distribution of reflectance values of each sub–target showed strong evidence for being normally distributed. Evidence for this is provided in table 5.15 (for BC) and table 5.16 (for DC). These tables show the summary statistics for the two sub–targets.

Once the concrete target was segmented the sample variograms for the two sub–targets were computed. Directional and omni–directional variograms for DC are shown in figure 5.16. The directional variograms showed weak evidence for a directional component, hence this was ignored. For BC (figure 5.16) there were only 87 measurements, providing insufficient data to test for a directional component, which was, therefore, ignored.

For DC, given the shape of the sample variogram, the circular, spherical and exponential models were fitted. Descriptive statistics are shown in table 5.17. Two procedures were

Waveband	Minimum	Median	Mean	Maximum	St. Dev	Skewness	Kurtosis
Band 1	12.36	16.73	16.55	19.77	1.85	-0.24	-0.91
Band 2	17.08	21.09	21.33	25.26	2.01	0.05	-1.00
Band 3	19.21	23.54	23.86	28.50	2.30	0.15	-1.07
Band 4	24.18	28.10	28.50	32.64	2.10	0.27	-1.04

TABLE 5.15: Summary statistics for the MMR data gathered over the bright section of the concrete surface (BC).



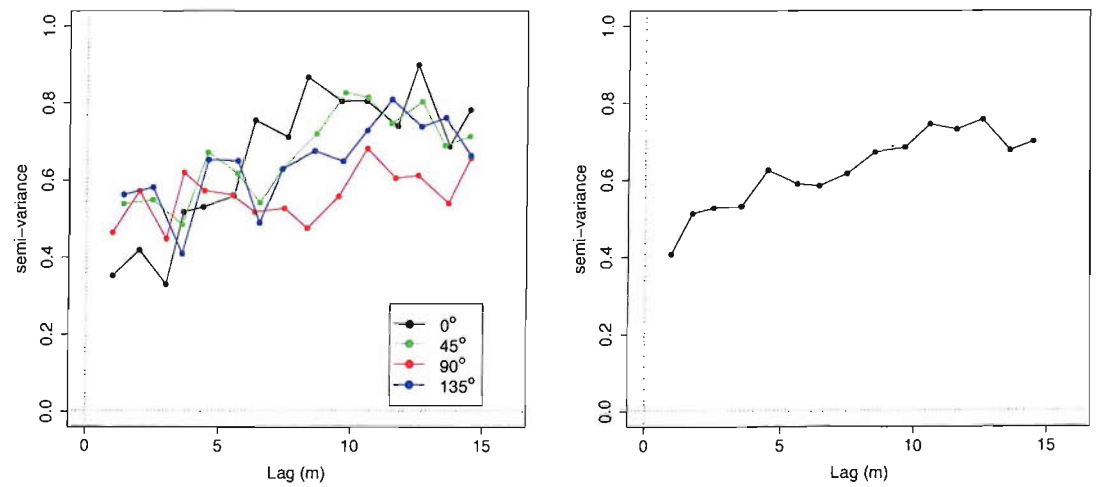
(a) Concrete surface (unlabelled slabs).

(b) Concrete surface (labelled slabs).

FIGURE 5.15: Figure showing the reflectance measurements made over the concrete surface, together with the boundaries and labels for the slabs. The size of the point is proportional to the magnitude of the reflectance measurement.

Waveband	Minimum	Median	Mean	Maximum	St. Dev	Skewness	Kurtosis
Band 1	11.65	13.76	13.71	15.46	0.83	-0.11	-0.40
Band 2	16.27	19.08	19.18	22.32	1.32	0.23	-0.53
Band 3	18.09	21.70	21.78	25.28	1.81	0.20	-0.04
Band 4	22.86	27.43	27.57	31.81	1.72	0.003	-0.44

TABLE 5.16: Summary statistics for the MMR data gathered over the dark section of the concrete surface (DC).



(a) Directional variogram

(b) Omnidirectional variogram

FIGURE 5.16: Directional and omnidirectional sample variograms for MMR band 1 over the dark concrete surface (DC).

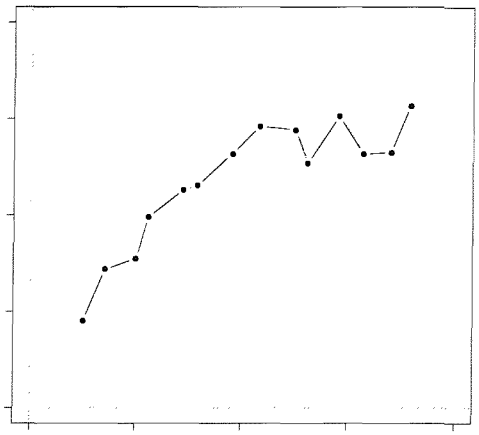


FIGURE 5.17: Omnidirectional sample variogram for MMR band 1 over the bright concrete surface (BC).

used for OK. In the first case, kriging was conducted using the global neighbourhood (GN). In the second, it was conducted using a local neighbourhood defined by the points contained in the spatially coincident slab (SN). The cross–validation diagnostics showed that defining the local neighbourhood on this basis increased the accuracy of prediction for all three models. There was little basis on which to choose between the three models and all gave a low value for the slope and R^2 value. The exponential model was chosen and kriging conducted in the local "slab" neighbourhood (SN). In conducting this, however, it should be noted that the "slope" and "R^2" shows that the variance explained by kriging was low (approximately 10%).

The results for the parameter estimation under the MBG–ML approach are shown in table 5.18. The results from the cross validation were similar to when the conventional approach was adopted. Using the exponential model resulted in the lowest value for the AIC and the highest value for the slope and R^2. On this basis, the exponential model was chosen and kriging conducted in the local "slab" neighbourhood. In conducting this, however, it should be noted that the "slope" and "R^2" shows that the variation explained by kriging was low (approximately 10%).

For BC, given the shape of the sample variogram, the circular, spherical and exponential models were fitted. The exponential model was rejected because it led to a substantially less accurate fit. Descriptive statistics are shown in table 5.19. Two procedures were used for OK. In the first case, kriging was conducted using the global neighbourhood. In the second, it was conducted using a local neighbourhood defined by the points contained in the spatially coincident slab. The cross–validation diagnostics showed that defining the local neighbourhood on this basis increased the accuracy of prediction for both models. However, the reduction in the MSDR when kriging was conducted in the local "slab" neighbourhood implied that the MSE underestimated accuracy, relative to the kriging variance. There was little basis on which to choose between the spherical and circu-

Model	Variogram				Cross–validation				
	τ_n^2	τ_s^2	ϕ	SS	ME	MSE	MSDR	*Slope*	R^2
Circular (GN)	0.44	0.28	12.20	7.33	0.002	0.63	1.16	0.09	0.07
Circular (SN)					-0.002	0.56	1.02	0.13	0.08
Spherical (GN)	0.43	0.29	12.90	7.88	0.003	0.63	1.17	0.09	0.07
Spherical (SN)					-0.002	0.56	1.03	0.13	0.08
Exponential (GN)	0.40	0.38	7.28	8.95	0.002	0.62	1.15	0.10	0.09
Exponential (SN)					-0.002	0.56	1.03	0.14	0.09

TABLE 5.17: Comparison of different variogram models fitted to MMR band 1 over the concrete surface (DC). The variogram model was fitted to the sample variogram and cross–validation prediction conducted using OK. GN refers to conducting OK in the global neighbourhood and SN refers to conducting OK in the local neighbourhood defined by the slab.

Model	Variogram				Cross–validation				
	τ_n^2	τ_s^2	ϕ	AIC	ME	MSE	MSDR	*Slope*	R^2
Circular (GN)	0.53	0.25	11.7	403	0.002	0.63	1.01	0.07	0.07
Circular (SN)					-0.001	0.57	0.88	0.12	0.08
Spherical (GN)	0.52	0.27	12.8	403	0.003	0.63	1.01	0.08	0.07
Spherical (SN)					-0.002	0.57	0.88	0.12	0.08
Exponential (GN)	0.22	0.54	1.77	399	0.006	0.61	1.01	0.10	0.10
Exponential (SN)					-0.001	0.55	0.94	0.16	0.11

TABLE 5.18: Comparison of different variogram models fitted to MMR band 1 over the concrete surface (DC). The models were fitted using ML and cross–validation prediction conducted using OK. GN refers to conducting OK in the global neighbourhood and SN refers to conducting OK in the local neighbourhood defined by the slab.

lar models. The spherical model was chosen and kriging conducted in the local "slab" neighbourhood.

The results for the parameter estimation under the MBG–ML approach are shown in table 5.20. The results from the cross validation are similar to when the conventional approach was adopted. There is little basis on which to choose between the three models since the difference in the AIC and cross–validation diagnostics for the different models is small. The spherical model was chosen and kriging conducted in the local "slab" neighbourhood.

5.4.2.3 MMR band 1: grass GT

Directional and omnidirectional variograms are shown in figure 5.18. Together with the spatial summary plot given in figure 4.7(a), these demonstrated the presence of spatial structure. The omnidirectional variogram shows that the nature of the spatial variation was broadly consistent in all directions, although there is some indication of structural anisotropy. However, estimating an anisotropy parameter within the conventional geostatistical framework is not straightforward and is generally performed by "trial and error",

Model	Variogram				Cross–validation				
	τ_n^2	τ_s^2	ϕ	SS	ME	MSE	MSDR	*Slope*	R^2
Circular (GN)	0.24	2.62	4.45	54	-0.03	1.33	1.01	0.55	0.56
Circular (SN)					-0.02	0.80	0.60	0.75	0.74
Spherical (GN)	0.10	2.76	4.88	55	-0.03	1.26	0.96	0.56	0.56
Spherical (SN)					-0.02	0.80	0.63	0.76	0.74

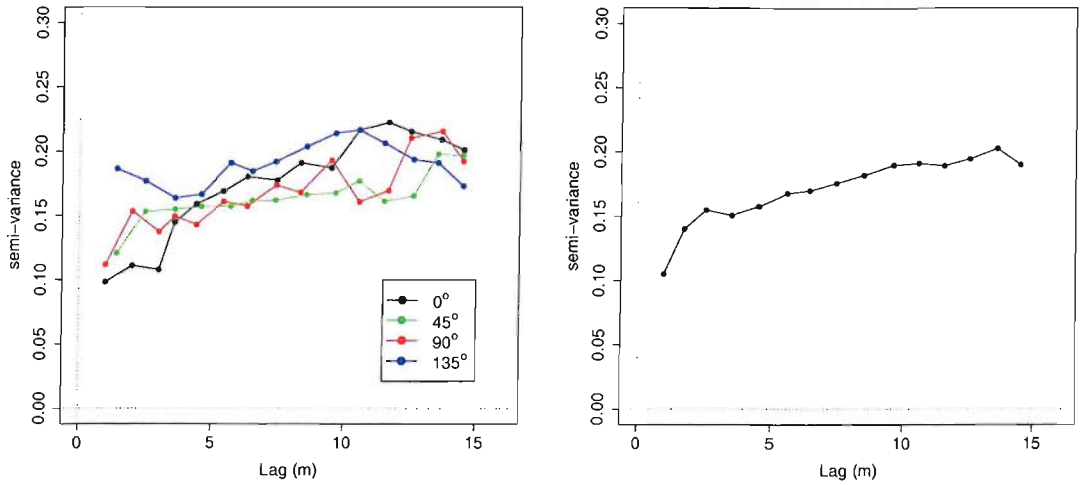
TABLE 5.19: Comparison of different variogram models fitted to MMR band 1 over the concrete surface (BC). The variogram model was fitted to the sample variogram and cross–validation prediction conducted using OK. GN refers to conducting OK in the global neighbourhood and SN refers to conducting OK in the local neighbourhood defined by the slab.

Model	Variogram				Cross–validation				
	τ_n^2	τ_s^2	ϕ	AIC	ME	MSE	MSDR	*Slope*	R^2
Circular (GN)	0	3.89	5.49	290	-0.02	1.24	1.04	0.61	0.59
Circular (SN)					-0.02	0.75	0.65	0.77	0.75
Spherical (GN)	0	3.58	5.93	292	-0.02	1.23	0.99	0.60	0.60
Spherical (SN)					-0.02	0.76	0.66	0.77	0.75
Exponential (GN)	0	3.99	3.88	294	-0.01	1.21	0.93	0.58	0.60
Exponential (SN)					-0.02	0.78	0.64	0.75	0.75

TABLE 5.20: Comparison of different variogram models fitted to MMR band 1 over the concrete surface (BC). The parameters were estimated using ML and cross–validation prediction conducted using OK. GN refers to conducting OK in the global neighbourhood and SN refers to conducting OK in the local neighbourhood defined by the slab.

rather than formal estimation (Goovaerts 1997, Pebesma & Wesseling 1998, Pebesma 2003). Hence, other approaches were considered. The MBG–ML approach is sufficiently flexible to allow the anisotropy parameters to be added to the likelihood function and estimated by maximum likelihood. A further possible approach is to fit a nested model to the omnidirectional sample variogram. Under the conventional approach, this is straightforward to do and can be implemented in `gstat`. Under the MBG–ML approach, the likelihood function can, theoretically, be extended in this way (Diggle et al. 2003), although they also suggest that the data are unlikely to support such models. It may also be noted that, in this case, under MBG–ML including anisotropy parameters may be used to increase the accuracy of estimation and prediction.

Under the conventional approach four options were considered. These were to: (i) fit a non–nested model and ignore the trend (table 5.21); (ii) fit a non–nested model and include the trend (table 5.22); (iii) fit a nested model and ignore the trend (table 5.23); and (iv) to fit a nested model with a trend (table 5.24). The cross–validation diagnostics suggested that the nested model was most accurate. However, they also suggested that, when a nested model was fitted, no increase in accuracy was gained by modelling the trend. On this basis, the nested model without a trend was chosen. Using the double exponential model led to numerical instability in the fitting procedure, leaving a choice between the double circular and double spherical models. There was little evidence to



(a) Directional variogram

(b) Omnidirectional variogram

FIGURE 5.18: Omnidirectional and directional sample variograms for MMR band 1 over the grass surface.

	Variogram				Cross–validation				
Model	τ_n^2	τ_s^2	ϕ	SS	ME	MSE	MSDR	*Slope*	R^2
Spherical	0.13	0.07	11.48	0.43	0.005	0.15	0.99	0.21	0.21
Circular	0.12	0.07	13.26	0.41	0.005	0.15	0.98	0.21	0.21
Exponential	0.11	0.09	6.05	0.42	0.003	0.14	0.96	0.22	0.22

TABLE 5.21: Comparison of different variogram models fitted to MMR band 1 over the grass surface. The variogram model was fitted to the sample variogram and cross–validation prediction conducted using OK.

choose between the double–spherical and double–circular models. The double–spherical model was chosen.

When the MBG–ML approach was adopted it was necessary to take account of the trend when estimating the variogram. This is because the MBG–ML approach fits the model to all the data, as one does not restrict the maximum lag. Note that, if the anisotropy parameters are estimated, this will alter the nature of the trend. Table 5.25 shows the results for the most simple case (no trend and no anisotropy parameters). Tables 5.26

	Variogram				Cross–validation				
Model	τ_n^2	τ_s^2	ϕ	SS	ME	MSE	MSDR	*Slope*	R^2
Circular	0.13	0.05	9.86	0.51	-0.01	0.14	0.97	0.21	0.21
Spherical	0.13	0.06	10.79	0.51	-0.05	0.14	0.97	0.21	0.21
Exponential	0.10	0.08	3.61	0.55	-0.01	0.14	0.96	0.23	0.24

TABLE 5.22: Comparison of different variogram models fitted to MMR band 1 over the grass surface. The variogram model is fitted to the sample variogram after first fitting a first–order trend on location. Cross–validation prediction was conducted using KT.

Model	Variogram						Cross–validation				
	τ_n^2	$\tau_{s_1}^2$	ϕ_1	$\tau_{s_2}^2$	ϕ_2	SS	ME	MSE	MSDR	*Slope*	R^2
Double Circular	0.05	0.09	1.99	0.06	12.53	0.22	0.005	0.13	0.94	0.30	0.33
Double Spherical	0.03	0.1	2.03	0.07	13.96	0.22	-0.007	0.11	0.91	0.30	0.34

TABLE 5.23: Comparison of different nested variogram models fitted to MMR band 1 over the grass surface. The variogram model was fitted to the sample variogram and cross–validation prediction conducted using OK.

Model	Variogram						Cross–validation				
	τ_n^2	$\tau_{s_1}^2$	ϕ_1	$\tau_{s_2}^2$	ϕ_2	SS	ME	MSE	MSDR	*Slope*	R^2
Double Circular	0.05	0.09	2.00	0.05	10.42	0.30	-0.01	0.12	0.92	0.30	0.32
Double Spherical	0.03	0.1	2.07	0.05	11.67	0.33	-0.01	0.12	0.91	0.30	0.33

TABLE 5.24: Comparison of different nested variogram models fitted to MMR band 1 over the grass surface. The variogram model is fitted to the sample variogram after first fitting a first–order trend on location. Cross–validation prediction was conducted using KT.

and 5.27 show the cases when the trend or when the anisotropy parameters were estimated, respectively. Both of these models had a lower AIC and more accurate cross–validation accuracy than the simple case. The models that included the anisotropy parameters had a slightly higher AIC than those that only included the trend (suggesting a less accurate fit). However, the cross–validation diagnostics suggested that the models including the anisotropy parameters were more accurate. When the trend *and* the anisotropy parameters were both estimated (table 5.28), the AIC decreased further. This suggested that this model fitted the data most accurately. However, the cross–validation diagnostics suggest that this model was no more accurate than the model that *only* included the anisotropy parameters.

The above discussion emphasised that the choice of models under the MBG–ML approach was not clear cut. The most complicated model (i.e., the circular covariance structure incorporating the trend and anisotropy parameters) resulted in the lowest AIC and highest values for the slope and R^2. This model was adopted.

Model	Variogram				Cross–validation				
	τ_n^2	τ_s^2	ϕ	AIC	ME	MSE	MSDR	*Slope*	R^2
Spherical	0.11	0.12	16	249.7	-0.01	0.14	0.99	0.23	0.23
Circular	0.12	0.09	10.66	253.6	-0.01	0.14	0.98	0.22	0.21
Exponential	0	0.23	1.83	248.6	-0.01	0.12	0.96	0.31	0.34

TABLE 5.25: Comparison of different variogram models fitted to MMR band 1 over the grass surface. The model was fitted using ML. Cross–validation prediction was conducted using OK.

Model	Variogram				Cross–validation				
	τ_n^2	τ_s^2	ϕ	AIC	ME	MSE	MSDR	*Slope*	R^2
Spherical	0	0.17	2.3	236.6	-0.03	0.12	0.96	0.30	0.32
Circular	0.04	0.14	2.5	238.1	-0.03	0.12	0.97	0.29	0.31
Exponential	0	0.18	1.23	234.7	-0.02	0.12	0.94	0.30	0.35

TABLE 5.26: Comparison of different variogram models fitted to MMR band 1 over the grass surface. The model was fitted using ML, including a first order trend on location. Cross–validation prediction was conducted using KT.

Model	Variogram						Cross–validation				
	τ_n^2	τ_s^2	ϕ	ψ_A	ψ_R	AIC	ME	MSE	MSDR	*Slope*	R^2
Spherical	0.07	0.12	2.88	0.04	4.86	240.5	-0.01	0.11	1.09	0.35	0.36
Circular	0.06	0.19	5.68	0.09	2.48	238.2	-0.01	0.11	1.06	0.37	0.39
Exponential	0.06	0.15	2.43	3.23	3.09	241.9	-0.12	0.12	1.12	0.33	0.35

TABLE 5.27: Comparison of different variogram models fitted to MMR band 1 over the grass surface. The model was fitted using ML, including the anisotropy parameters. Cross–validation prediction was conducted using KT.

5.4.2.4 Summary of the geostatistical analysis

The geostatistical analysis clearly illustrated spatial dependence in the reflectance fields for the three surfaces. However, the cross validation results were quite disappointing, particularly for the low–reflectance concrete surface (DC).

Conventional geostatistical and MBG–ML approaches were adopted. The differing approaches have developed from different schools of thought. "Conventional" geostatistics was developed, primarily, by quantitative geologists and soils scientists. Those scientists and engineers needed to solve practical problems and needed workable solutions and have a well documented history of success. The MBG approach has been developed mainly by statisticians who have emphasised theoretical considerations. However, it should be noted this development has been considerably reinforced by practical considerations.

Given the different implementations, it is reassuring to note that the two approaches give comparable results, as illustrated by the accuracy assessment. This is the case, even if the estimated parameters of the variogram differ. The conventional approach requires pre–

Model	Variogram						Cross–validation				
	τ_n^2	τ_s^2	ϕ	ψ_A	ψ_R	AIC	ME	MSE	MSDR	*Slope*	R^2
Spherical	0.07	0.09	2.51	3.2	5.02	224.2	-0.02	0.11	1.06	0.35	0.37
Circular	0.07	0.09	2.18	3.2	5.12	223.4	-0.02	0.11	1.06	0.36	0.37
Exponential	0.06	0.09	0.82	0.05	5.46	230.6	-0.02	0.12	1.06	0.33	0.34

TABLE 5.28: Comparison of different variogram models fitted to MMR band 1 over the grass surface. The model was fitted using ML, including a first order trend on location and the anisotropy parameters. Cross–validation prediction was conducted using KT.

processing to calculated the sample variogram as well as expert judgement when deciding what model to fit. The MBG–ML approach is attractive in this respect since there is less pre–processing and the method favours automation. Exploratory data analysis is valuable to the user to assess the problem. However, the subjective judgement that would have taken place in the pre–processing and variography is explicitly accounted for in the model formulation. Choices between models can be made on the basis of model accuracy and the AIC.

When adopting the conventional approach the following models were used:

- Asphalt: Circular model with no trend. 16 m neighbourhood.
- Concrete (DC): Exponential model, no trend. Local neighbourhood defined by slab.
- Concrete (BC): Spherical model, no trend. Local neighbourhood defined by slab.
- Grass: Double spherical model, no trend. 16 m neighbourhood.

When adopting the MBG–ML approach the following models were used:

- Asphalt: Circular model with a first order trend on location. Global neighbourhood.
- Concrete (DC): Exponential model, no trend. Local neighbourhood defined by slab.
- Concrete (BC): Spherical model, no trend. Local neighbourhood defined by slab.
- Grass: Circular model, incorporating anisotropy and trend. Global neighbourhood.

5.4.2.5 Implementing the regression model

The variogram models were selected for each surface, as discussed in the previous section. The locations of pixels lying within the targets were identified and block kriging was used to predict the value of reflectance for spatially coincident blocks, defined on the same support and at the same locations as the pixels. Prediction was conducted first using block–kriging and then using block–conditional–simulation. The results presented here are for the 2 m pixels. Kriging and conditional simulation were conducted using the variogram parameters calculated using "conventional" geostatistics. Comparable results were obtained when the MBG-ML estimated variogram parameters were used – hence the results are not presented here.

For each target, the pixels that overlapped, or bordered on the edge of the target were omitted (to limit edge effects). The remaining remotely sensed pixel values and predicted (block–kriged) reflectance values were combined and used to estimate the parameters of the regression model. This resulted in 368 data pairs being available for the regression model.

Parameter	Estimate	Std. Error	p value
β_0	-2.3882	0.0711	< 0.01
β_1	0.3005	0.0019	< 0.01
σ^2	0.2592		
R^2	0.98		

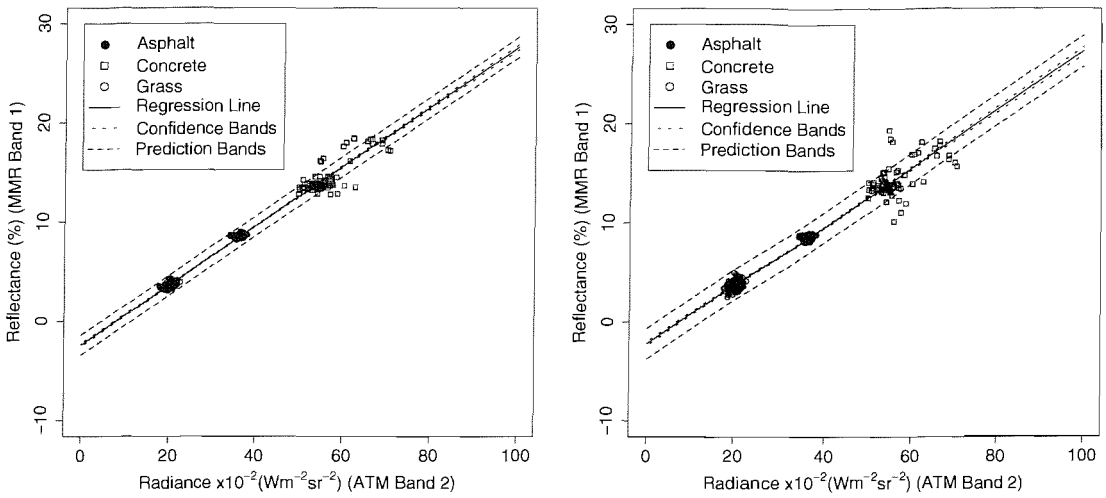
TABLE 5.29: Table showing the results of the parameter estimation for the ELM regression model after the reflectance measurements have been aggregated using block–kriging. The estimates are for MMR band 1 regressed on ATM band 2.

The resulting parameter estimates for the kriged blocks are shown in table 5.29 and the resulting regression is illustrated in figure 5.19(a). It is clear from figure 5.19(a) that the between-target variance was much greater than the within-target variance. This effect was also observed for the point-pixel approach. This is for two reasons. First, the kriging cross–validation outlined in the previous sections showed that, for all targets the accuracy of prediction was low. Second, kriging smoothes.

By comparison to the typical approach it was observed that the estimate of β_0 was 0.4% (reflectance) lower and the estimate of β_1 was slightly lower (0.3005 vs 0.3118). However, the results imply substantially more confidence in this regression since $p < 0.01$ for both parameter estimates, which was not the case under the typical approach. Of particular interest, from the perspective of uncertainty analysis, was that the estimate of σ^2 is two orders of magnitude larger (0.26 vs 0.0017). This reflected the fact that the data are defined on different supports between the two approaches. Under the typical approach the field and remotely sensed data were averaged for each target (approximately 32 m $\times$ 32 m). For the block-pixel approach the support of the reflectance and radiance data reflected the size of the pixels in the image that required atmospheric correction.

By comparison to the point-pixel approach, it was observed that the estimates of β_0 (-2.39 vs -2.42) and β_1 (0.3008 vs 0.3005) were similar. However, the estimate of σ^2 was smaller (0.26 vs 0.78). This reflects, in part, the fact the reflectance data were defined on different supports between the two approaches. Part of this reduction was also attributed to the smoothing effect of kriging, which is discussed below. Under the point-pixel approach, the data were defined on pseudo–point supports, although the radiance data were defined on pixel–sized supports. By aggregating from the point to block support, the variance in the reflectance data (the y_i) was reduced. This had implications for the calculation of confidence and prediction intervals, which tend to be narrower than for the point-pixel approach.

Recall from section 2.2.2.6 that kriging is the *Best Linear Unbiased Predictor* (BLUP). Hence, kriging *smoothes* with the result that variation is reduced and a kriged surface is *not* a possible reality. This has important implications for uncertainty analysis. In order to generate possible realities, block conditional simulation (sequential Gaussian simulation (SGS)) was used to produce 1000 simulations for each target. Each of these possible realities was regressed against the spatially coincident radiance data. The variation in the



(a) Regression for the kriged blocks.

(b) Regression for one realisation (Sim516).

FIGURE 5.19: Plots showing regressions of (a) the kriged blocks against the remotely sensed pixels and (b) conditionally simulated blocks against the remotely sensed pixels for MMR band 1 and ATM band 2.

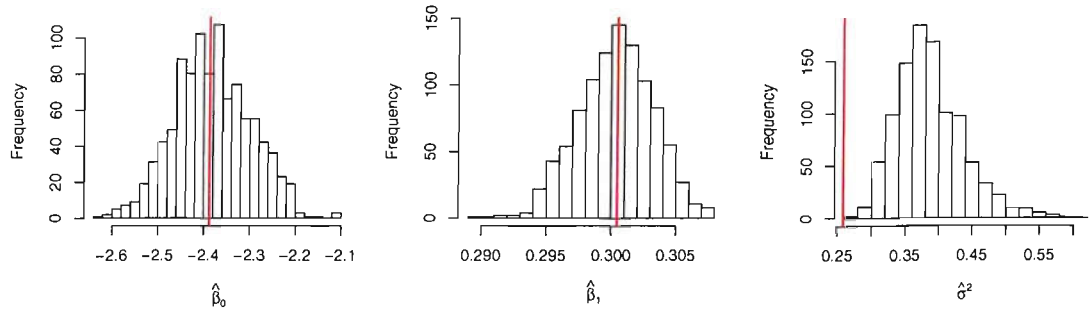
Parameter	Mean	Median	Std. Dev.	Skewness	Kurtosis
β_0	-2.383	-2.381	0.0841	0.10	-0.14
β_1	0.3005	0.3005	0.0028	-0.14	-0.11
σ	0.6167	0.6208	0.0505	0.79	1.02

TABLE 5.30: Table showing the summary statistics for the estimated parameter values obtained from the simulated surfaces. The estimates are for MMR band 1 regressed on ATM band 2.

estimates of β_0, β_1 and σ^2 are illustrated in table 5.30 and figure 5.20. The estimates of β_0 and β_1 for the conditionally simulated surfaces varied about a mean which corresponded to the the estimate obtained from the kriged surfaces. However, the estimates of σ^2 were all larger than the estimate gained for the kriged surface. This reflected the smoothing in the kriged surface and the result was that the confidence and prediction intervals were increased. This is clearly illustrated by comparing figures 5.19(a) and 5.19(b)

The implications of this for prediction are illustrated at two x–values, $x = 40$ and $x = 50$. This is shown in tables 5.31 and 5.32. This clearly demonstrates that using the kriged surface for prediction led to a narrowing of the prediction interval by 1% reflectance at $x = 40$, by comparison to the largest estimated value of σ^2 (at Sim516) from the conditionally simulated surfaces. This example illustrated that using the block kriged surface underestimates σ^2 leading to an underestimation of the confidence and prediction intervals. Hence, if geostatistics are used to conduct the aggregation it is necessary to analyse the results from the conditional simulation in order to ensure realistic assessment of the uncertainty in estimation and prediction.

The above analysis demonstrated the problem with using kriged surfaces in regression.



(a) Histogram for $\hat{\beta}_0$ (b) Histogram for $\hat{\beta}_1$ (c) Histogram for $\hat{\sigma}^2$

FIGURE 5.20: Histograms showing the distribution of the parameter estimates for the 1000 realisations of the simulated surfaces. The red line shows the estimated parameter value for the kriged surfaces (note that this lies outside the range for $\hat{\sigma}^2$).

Simulation	0.05%	2.5%	mean	97.5%	99.5%
Kriged ($\hat{\sigma}^2 = 0.26$)	8.31	8.63	9.63	10.63	10.95
Sim1 ($\hat{\sigma}^2 = 0.37$)	8.04	8.42	9.63	10.83	11.21
Sim516 ($\hat{\sigma}^2 = 0.60$)	7.58	8.06	9.59	11.12	11.60

TABLE 5.31: Table showing the predicted value of reflectance together with the 95% and 99% prediction intervals for a radiance value of $x = 40 \times 10^2$ W m^{-2} sr^{-1}. The reflectance data are provided by the kriged blocks (first row) and two realisations of conditionally simulated blocks (second and third rows).

Kriging produced a smoothed surface. Hence the surface had a lower overall variance than a realisation of a conditionally simulated surface. If the kriged surface is then used in regression this may lead to biased estimates in β_0 and β_1, although in the case of the ELM this effect is mitigated since the between–GT variance is much larger than the within–GT variance. However, utilising the smoothed GTs *does* lead to bias (underestimation) of the residual variance, σ^2. Hence, the objective was to address this through conditional simulation, since conditionally simulated surfaces exhibit variance properties that better approximate reality. Nevertheless, σ^2 was generally lower than it was for the simple averaging approach ($\hat{\sigma}^2 = 0.4981$).

Simulation	0.05%	2.5%	mean	97.5%	99.5%
Kriged ($\hat{\sigma}^2 = 0.26$)	11.31	11.63	12.63	13.64	13.96
Sim1 ($\hat{\sigma}^2 = 0.37$)	12.61	11.41	12.61	13.82	14.20
Sim516 ($\hat{\sigma}^2 = 0.60$)	10.53	11.01	12.54	14.07	14.55

TABLE 5.32: Table showing the predicted value of reflectance together with the 95% and 99% prediction intervals for a radiance value of $x = 50 \times 10^2$ W m^{-2} sr^{-1}. The reflectance data are provided by the kriged blocks (first row) and two realisations of conditionally simulated blocks (second and third rows).

5.4.3 The block-pixel approach: summary and conclusions

The analysis conducted in this section sought to aggregate the field based measurements of reflectance to pixel-sized supports. The objective of this was to provide the two variables on equivalent spatial supports for input into the regression model. As discussed in section 2.2.1, this is more rigorous theoretically than combining data sets that are defined on different (and even unknown) spatial supports.

At the outset it was considered that geostatistics offered the optimal and most elegant solution to aggregation from point to block supports. This is because geostatistics makes explicit use of a spatial model and is well suited to use with a spatially structured sample (Brus & de Gruijter 1997, Bierkens et al. 2000). In contrast design-based approaches are based on a random sample and tend to require a larger sample size (i.e., less efficient), since understanding of the spatial structure is not incorporated into the sampling strategy.

However, upon implementation it was found that the results of the geostatistical analysis were disappointing. Despite the clear evidence of spatial structure demonstrated in this and previous studies (Salvatori et al. 2000), the accuracy of the kriged predictions was low. Block kriging was not fully presented for the other MMR wavebands, although more accurate results were obtained for the grass and concrete targets in MMR band 4. Furthermore, preliminary results on a transect of measurements suggested that accuracy can be increased if the minimum sampling interval is reduced from 1 m to 0.5 m. However, increasing the density of the sampling grid to 0.5 m intervals increases the sample size by a factor of four, thus increasing the length of time required to sample a target.

There are further theoretical and practical problems with using the geostatistical approach. Since kriging smoothes the kriged surfaces are not well suited for subsequent use in the regression model. It was also noted in section 2.2.2.6 that conditionally simulated surfaces are not ideal either since the simulation "adds back" variation that is orthogonal to the primary variable. Hence, it has been recommended that conditionally simulated results should not be used in regression (Curran & Atkinson 1998, Atkinson & Kelly 1997, Atkinson 1999b). Atkinson & Kelly (1997) and Atkinson & Tate (2000) suggest using an adjustment factor, but their results are not conclusive. In addition to these issues it is not clear how the multiple outcomes of the regression (associated with multiple realisations of conditionally simulated) should be interpreted.

Further to these theoretical issues, there is a practical problem with implementation of a geostatistical approach to aggregation. Implementation requires some training in, and understanding of, geostatistics. Furthermore, the implementation can be time-consuming. It is likely, therefore, that many users of remotely sensed data will not wish to devote time and effort to this unless the motivation is compelling (Atkinson 2005).

Despite the above mentioned problems, it should be recognised that geostatistics does provide a theoretically well recognised approach to aggregation from point to block supports. Hence, it does address the problem that the field and remotely sensed data were

defined on different spatial supports, albeit with the above caveats. Furthermore, it will be demonstrated in chapter 6 that this approach is not sensitive to reasonable positional uncertainty in the location of the field measurements.

In contrast to the geostatistical approach the simple averaging approach was straight forward to implement. It yielded point estimates for β_0 and β_1 that were consistent with those obtained for block-kriging and the point-pixel approach. However, the estimate of σ^2 is more appropriate since it lies between that obtained for kriging (too smooth) and that obtained for the point-pixel approach (too rough). The process of block kriging incorporates observations from outside the block (see figure 2.4) that are weighted by the kriging weights. This effectively increases the support of the block beyond its notional boundaries. This effect does not exist for simple averaging. Hence it is expected that kriged blocks will be smoother than simple–averaged blocks. As with the geostatistical approach this approach could be improved by obtaining a denser sample. However, obtaining a random sample would increase the complexity and time taken to sample each GT. Nevertheless, if the user is prepared to accept the drawbacks of applying the simple averaging approach then this approach bears considerable merit.

5.5 Summary and advice for the practitioner

This chapter began with a detailed analysis of the typical approach to implementation of the ELM. It was demonstrated that implementing the typical approach is problematic since averaging over blocks reduces the number of degrees of freedom in the regression to a minimum. Furthermore, averaging over GTs leads to a reduction in variance, leading to an unexpectedly low and unstable estimate of σ^2. These factors combine to give wide confidence intervals on the parameter estimates of β_0 and β_1 and on the regression line as a whole. In some cases, it also leads to the conclusion that the parameter estimates are not *statistically* different from zero. This latter conclusion is incorrect, since it is known from theory that $\beta_0 < 0$ and $\beta_1 > 0$. However, this result is correctly indicative of high uncertainty in the location of the regression line.

A further problem with the typical approach became clear when the sample size was reduced. This led to substantial variability in the parameter estimates. This is exacerbated by the fact that the field–based reflectance measurements suggest that there are two "sub-targets" within the concrete GT. It might be argued that this is a limitation of the GT *rather than* the typical approach and such a target would, ideally, not be used. However, this would leave the user without a bright target in visible wavebands. It should be recognised that the UK is not well endowed with "ideal" GTs. Hence a legitimate question is whether the ELM can be adapted to work with the type of targets available.

The modified-typical approach was adopted as a solution to the above issue. This split the concrete GT into two sub-targets. It should be recognised that this approach is only possible because the location of the field measurements were recorded.

Since the location of the field measurements was recorded, it was argued that better use could be made of the field-based measurements of reflectance. Two categories of approach were proposed: point-pixel and block-pixel.

For the point-pixel approach, the field-measurements of reflectance were combined directly with the pixel-based measurements of radiance. This gave substantially greater confidence in the ELM regression as demonstrated by the narrow confidence intervals on the parameters and the regression line. However, the prediction intervals were wide, as expected given the variance of the residuals. At least 50 (and ideally 100) measurements should be taken per-target to ensure accurate estimation of σ^2 and the associated confidence and prediction intervals.

The point-pixel approach makes better use of the data than the typical approach. It does not "lose the knowledge" contained in the field-data or lead to a dramatic reduction in the degrees of freedom. dramatic reduction in the degrees of freedom. It also allows greater freedom in the choice of GTs since the requirements apply at the pixel- rather than GT-scale. However, the point-pixel approach does not attempt to match the support of the field and remotely sensed data. Hence the estimated value of σ^2 will be too large.

To address the support issue two methods were investigated for aggregating from point to block supports. These were the simple averaging approach of averaging over all measurements in the pixel and the geostatistical approach of block kriging. It was expected that the block kriging should be the most suitable approach. However, it provided disappointing results since the kriged predictions were less accurate than hoped. Furthermore, there are theoretical and practical problems with using kriged or conditionally simulated data in a regression model. Finally, the geostatistical approach requires substantial expertise and may receive low take-up for this reason. Hence the simple averaging approach is likely to be preferable, if the user accepts the drawback that true design-based models should, theoretically, be implemented using randomly located data.

In terms of the estimation of β_0 and β_1, the point estimates were similar for the typical and modified-typical approaches. The estimates were also similar for the point-pixel and block-pixel approaches. However, for the latter approaches the estimates of β_0 were higher and estimates of β_1 lower than for the typical approach. As expected, the estimates of σ^2 were largest for the point-pixel approach and smallest for the typical approach – with the simple averaging block-pixel approach giving the most reliable estimates.

It is useful to give the confidence and prediction intervals further consideration. Confidence intervals quantify uncertainty in parameter estimates (or the position of the regression line), whereas prediction intervals quantify uncertainty in prediction of a "new" value of y. If the data show clear evidence for a trend, then it would be expected that the confidence intervals would be narrow. However, the prediction intervals are primarily determined by the residual variance. Hence, if the evidence for a trend is high but the residuals are also large then the confidence intervals may be narrow but the prediction intervals wide. This is shown in the point-pixel and block-pixel approaches. It is argued

that the ELM is a form of calibration hence, to explore variability in a predicted image of reflectance the regression line should be varied (within the confidence intervals). It does not make sense to draw individual pixels from within the prediction intervals since it is to be expected that the input image (radiance) should look the same as the predicted image of reflectance. However, this does not apply to all applications of empirical regression models. For example, if the objective was to predict biomass within a field (together with uncertainty estimates) then it would be appropriate to draw realisations from the predictive distribution. Indeed, in such a situation it would not be appropriate to draw each pixel within the field independently and, instead, conditional simulation should be used to model the field. This issue is returned to in chapter 7.

Given the wide availability of modern surveying and GPS equipment, recording the location of each field measurement is realistic for remote sensing research. The cost of such equipment is low by comparison to the overall cost of a remote sensing campaign, although it is recognised that additional trained personnel may be required if this is to be done simultaneously with measurement of reflectance. Chapter 6 addresses the possibility of uncertainty in location of the field measurements.

Finally it will be noted that, for the point-pixel and block-pixel approaches the regression residuals were shown to be heteroskedastic. Furthermore, since the targets are known to be characterised by spatial structure it is expected that the residuals from the regression may be auto-correlated. These two issues are addressed in chapter 7.

Chapter 6

Analysis II: positional uncertainty

In chapter 5 it was recommended that the accuracy and confidence in the ELM could be increased through utilisation of the knowledge of *location* of the field measurements. In the simplest case, this allowed splitting the sub-optimal concrete target into two more appropriate sub-targets. Further developments allowed implementation of the point-pixel (where field measurements are paired with their co-located pixel) and block-pixel (where the field-measurements are aggregated to pixel-sized blocks) approaches. Both these approaches allowed the development of more accurate regression models and allowed greater flexibility in the choice of GTs. The latter approach explicitly aims to estimate the ELM parameters on pixel-sized supports, which is the support required for prediction.

With the exception of the typical approach all subsequent analysis was contingent on knowing the location of the field measurements. Throughout the field work exceptional care was taken to record and subsequently survey the location of each field measurement. Hence the location of the field measurements was considered to be known with a high degree of confidence.

Modern surveying and differential GPS (DGPS) equipment allow the location of points on the Earth's surface to be recorded with a high degree of accuracy (< 0.01 m) (Letham 2003). However, it is recognised that it may not always be feasible for users to record location with such a high degree of accuracy. They may not have the necessary equipment or may be restricted by logistical or environmental considerations. However, even with an inexpensive hand-held GPS receiver (such as the Garmin eTrex[1]) horizontal location is recorded with a typical accuracy of less than 15 m (95% RMSE). This accuracy can be increased substantially through the use of DGPS, which is available even for these low cost instruments. For example, Garmin sell DGPS beacons for their receivers which enable the accuracy to be increased to 1–5 m (95% RMSE) in the horizontal domain. Furthermore the recently developed Satellite Based Augmentation System (SBAS) yields an accuracy of 3 m (95% RMSE) in the horizontal domain. SBAS is a recently developed system that relies on a network of ground-receiving stations. This uses a network of

[1]See: `www.garmin.com/etrex`

ground stations and designated satellites to gather and broadcast the required differential correction information. This can be interpreted by most recent hand held receivers, such as the Garmin eTrex. SBAS is known as the Wide Area Augmentation System (WAAS) in North America. The equivalent system in Europe is the European Geostationary Navigation Overlay Service (EGNOS) (Letham 2003).

The objective of the research undertaken for this chapter was to investigate the impact of uncertainty in the location of the field measurements on the output of the ELM. This analysis has broader significance in the application of empirical regression models in remote sensing. Section 6.3 is also relevant to the understanding of the impact of positional uncertainty on geostatistical modelling – a subject that has only been given limited attention in the literature. A general discussion of positional uncertainty is given in section 2.3.

For this section, analysis and discussion focuses on deriving broad-band reflectance for band 1 of the MMR (420–530 nm) using band 2 from the ATM (450–520 nm). As discussed in section 4.4 the location of each measurement was surveyed and recorded relative to UK Ordnance Survey (OS) trigonometric points. The ATM data were also geometrically corrected to the OS National Grid (RMSE < 1 pixel). This allowed each field measurement to be located within the image. Careful attention was given to ensuring that the locations of the field measurements were recorded rigorously and accurately, both relative to each other and to the OS National Grid. In an operational situation, it might not be possible to record location with such rigour and precision and this research seeks to examine the implication of that. However, the analysis assumes that the geometric correction of the imagery is perfect.

In examining the effect of positional uncertainty it was hypothesised that:

1. Introducing positional uncertainty will affect the estimate of β_0 and β_1, leading to a "flattening" of the regression line. Hence, the parameter estimates will be *biased* relative to the estimates made in the absence of positional uncertainty.

2. Introducing positional uncertainty will lead to a reduction in the accuracy of the parameter estimates and new variable prediction. This will be manifested by a positive bias in the estimate of σ^2 following the introduction of positional uncertainty.

3. Introducing positional uncertainty will lead to uncertainty in the parameter estimates, as manifested by variability in the parameter estimates between different realisations of positional uncertainty.

4. Decreasing the sample size will *increase* the effects of positional uncertainty – *i.e.*, the items listed in the first three hypotheses.

5. Increasing the pixel size will *reduce* the effect of positional uncertainty – *i.e.*, the items listed in the first three hypotheses.

6.1 Simulating positional uncertainty

In order to simulate positional uncertainty, a random error term, ε, was added to the Easting and Northing of each field location, as follows:

$$\begin{aligned} E_i &= \text{Easting} + \varepsilon_{x_i} \\ N_i &= \text{Northing} + \varepsilon_{y_i} \end{aligned} \tag{6.1}$$

where i refers to each individual field location. It was judged reasonable to model ε as being drawn from a Normal distribution, $\varepsilon \sim N(0, \cdot)$. Taking $\varepsilon \sim N(0, 0.25)$ reflects the case where the operator has a high degree of confidence in positional accuracy of the field measurements. If the user had not been able to give such attention to recording location, $\varepsilon \sim N(0, 4)$ is likely to be more realistic. This latter case is illustrated in figure 6.1. With these assumptions, different realisations of the sampling scheme for each surface can be simulated and used to explore the effect of positional uncertainty. These are termed the "perturbed" data sets and an example is given in figure 6.1. The original attribute was then assigned to the perturbed location. The effect of positional uncertainty was then analysed by reference to the resulting modelled variograms and the effect on the implementation of the ELM for both the point-pixel and block-pixel approaches.

Adopting this approach tackles the situation where a measurement is performed at the intended location but attributed to an incorrect location. This is the FP (resource) model (Gabrosek & Cressie 2002, Cressie & Kornak 2003) and discussed in section 2.3. It might be argued that the CP (design) model is more appropriate for this investigation. The basis for that argument would be that the locations were specified in advance and then realised imperfectly. Under the CP approach the attribute for the perturbed location would be assigned to the original location. However, the problem with this approach is that the attribute (reflectance) is unknown at the perturbed locations. Although this value could be simulated, the resulting attribute value would be uncertain and subsequent analysis would be several steps removed from the original data. It is recognised that the distinction between the CP and FP models is important when the objective is to correct for the effect of positional uncertainty (Cressie & Kornak 2003). However, the objective of the analysis undertaken in this chapter is to assess the impact of positional uncertainty, rather than to correct for it. Given this aim, it is argued that adopting the FP model is reasonable and is preferred to working within the CP framework and simulated attribute values.

6.2 Point-pixel

Each field measurement was linked with its spatially coincident pixel value, yielding a set of data pairs. These were used for parameter estimation in the ELM regression model. The full data set ($nFull$) comprised 696 measurements ($\sim$ 230 per site). This was sub-sampled

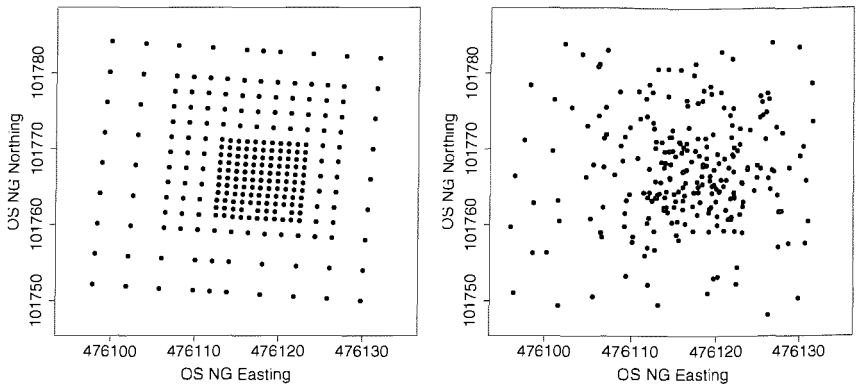


FIGURE 6.1: Plots showing the unperturbed and a realisation of the perturbed data set for $\varepsilon \sim N(0, 4)$.

to examine the impact of sample size on parameter estimation in the presence of positional uncertainty. Sub-samples of 100 ($n100$), 50 ($n50$), 30 ($n30$), 20 ($n20$), 10 ($n10$), 5 ($n5$), 3 ($n3$) and 1 ($n1$) measurements per site were created. Note that $n1 \subset n3 \subset \cdots \subset n100$.

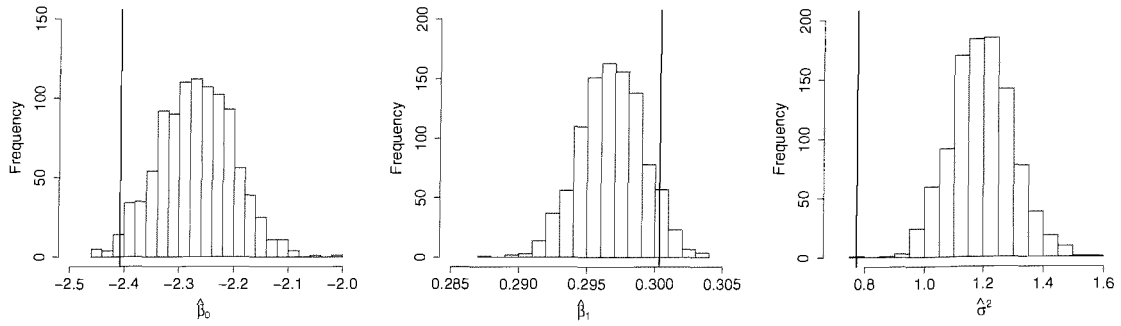
To examine the impact of pixel size, the image was degraded from 2 m pixels and the experiment repeated for 4 m, 6 m and 8 m pixel sizes.

In summary, the following approach was adopted:

1. For the full data set generate 1000 realisations of the field data, where each realisation differs only in that the location has been perturbed (see section 6.1).
2. Generate the sub-samples (see above).
3. Apply the same sub-samples to *each* realisation. Hence, for each realisation, the location will differ but the attribute will remain unchanged. This is done to isolate positional uncertainty from sampling variability (see section 6.2.2).
4. For each realisation, link the field measurement with its spatially coincident pixel measurement.
5. Apply the regression for each realisation and each sub-sample.
6. Repeat from step 1 for each pixel size.

6.2.1 Full data set

The first stage was to examine the impact of positional uncertainty on parameter estimation using the full data set and 2 m pixels. The result of estimating the parameters of the regression model for 1000 samples in the perturbed data set are shown in figure 6.2. This shows clearly that β_0 and σ^2 are overestimated relative to the estimate made in the absence of positional uncertainty. However, β_1 is underestimated. The increase in the



(a) Variation in $\hat{\beta}_0$. (b) Variation in $\hat{\beta}_1$. (c) Variation in $\hat{\sigma}^2$.

FIGURE 6.2: Plots showing how estimated parameters vary with the introduction of positional uncertainty, $\varepsilon \sim N(0, 4)$. The vertical line indicates the estimated parameter value in the absence of positional uncertainty.

	$\hat{\sigma}^2$	Prediction interval (%)
Unperturbed	0.77	3.4
2.5 percentile	0.99	3.9
Mean	1.19	4.3
Median	1.19	4.3
97.5 percentile	1.40	4.6

TABLE 6.1: Table showing how the introduction of positional uncertainty affects estimation of $\hat{\sigma}^2$ and the associated prediction intervals. Note that the width of the prediction intervals is approximately uniform ac cross the full range of the x-axis.

value of $\hat{\sigma}^2$ is indicative of the increase in uncertainty, both in parameter estimation and for prediction. It should be noted that the magnitude of the variability in $\hat{\beta}_0$ and $\hat{\beta}_1$ is small.

The results have an intuitive interpretation. It is reasonable to hypothesise that introducing positional uncertainty will decrease the correlation between the two variables. This will naturally lead to a "flattening" of the regression line.

The increase in $\hat{\sigma}^2$ is important to note because of its role in determining the confidence intervals and prediction intervals for predicting a new value. Summary results are shown in table 6.1 and in figure 6.3. For the unperturbed data set $\hat{\sigma}^2 = 0.77$ leading to a prediction interval of 3.4% reflectance. However, for the perturbed data set the prediction intervals are widened by between 0.5% and 1.2% reflectance (95th percentile). Note that the sample of $\hat{\sigma}^2$ obtained from the perturbed data set is approximately normally distributed (skewness = 0.23; kurtosis = 0.25) hence the median and mean value are approximately equal (two decimal places). Note also that strictly, the prediction intervals widen either side of the mean value on the x-axis. However, this is not observable to one decimal place hence a single value is quoted for the prediction interval for all values of x.

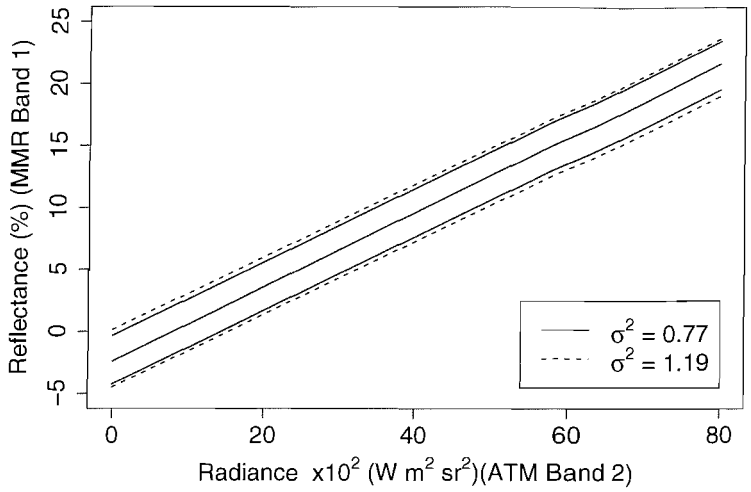
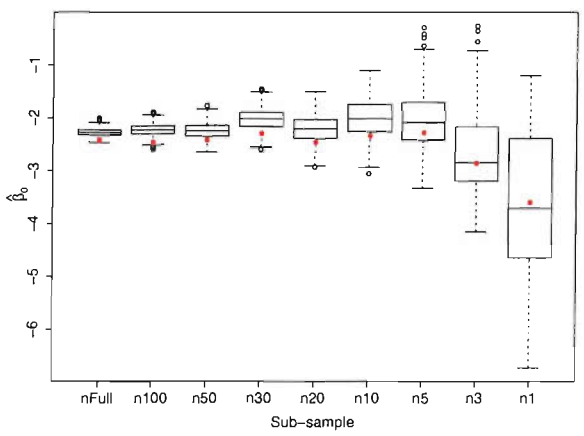


FIGURE 6.3: Plot showing how the increase in $\hat{\sigma}^2$ affects the prediction intervals around the regression line. Data points are not shown to improve clarity, but can be seen in figure 5.7(a)

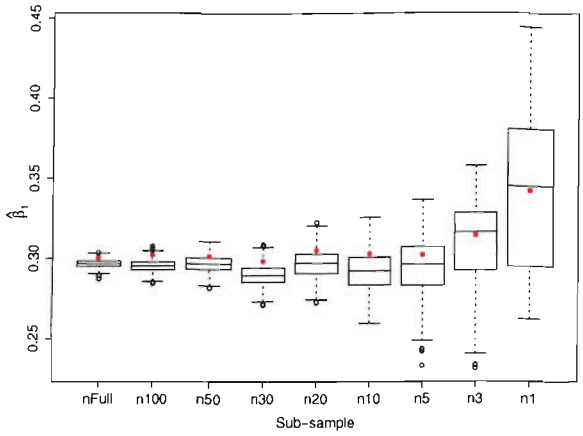
6.2.2 Reduced sample size

The next stage was to examine the impact of positional uncertainty for the sub-sampled data sets. This is illustrated in figure 6.4 for the 2 m pixels. It is clear that the variation in the parameter estimates, in the presence of positional uncertainty, increases as the sample size decreases. This is particularly apparent where less than 20 measurements are made per target. This figure of 20 measurements per target is suggested as an absolute *minimum* sample size in the absence of positional uncertainty (see section 5.3.2), although a sample size of 50 or more measurements per GT is recommended. Indeed, figure 6.4 shows that for $n50$ or larger the estimates of all three parameters were substantially less variable than for smaller sample sizes. Hence 50 is recommended as the minimum sample size under conditions of positional uncertainty. However, the user may need to further revise this number as discussed in section 6.2.4.

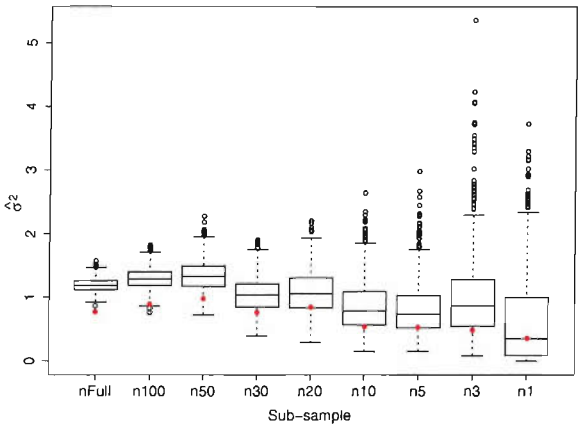
The data sets for the above analyses were generated from *one* realisation for each sub-sample. This means that for each of the 1000 perturbations, the same $n100$ samples were used. The experiment was designed in this way to isolate the effect of positional uncertainty from sample variability. The effect of sample variability for samples smaller than $nFull$ was discussed in section 5.3.1. It is unclear what the emergent effect of combining positional uncertainty and sample variability would be. It is postulated that for each sample size the *position* of the box-plot may be attributed to sample variability, although the shape of the box-plot is the result of uncertainty in the location of the field measurements. Hence, if sample variability and positional uncertainty were combined, the uncertainty in parameter estimation and prediction would be increased further. However, it may also be that positional uncertainty effectively smoothes the sample variability. The combined effects of positional uncertainty and sample size have not been studied for this



(a) Variation in $\hat{\beta}_0$.



(b) Variation in $\hat{\beta}_1$.



(c) Variation in $\hat{\sigma}^2$.

FIGURE 6.4: Plots showing how estimated parameters vary with the introduction of positional uncertainty, $\varepsilon \sim N(0, 4)$, for different sample sizes (point-pixel approach). The red circle indicates the estimated parameter value in the absence of positional uncertainty.

thesis and remain an avenue for future research. The implication is that a sample size larger than 50 per target may be advisable.

6.2.3 Pixel size

The next stage was to examine the impact of increasing the pixel size whilst holding the sample sized fixed. The effect is illustrated in Figures 6.5 to 6.7 for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ respectively. It will be noted that the estimate of β_0 and β_1, in the absence of positional uncertainty, changes slightly for different pixel sizes. This is a MAUP effect and was discussed in section 2.2.1.

There are two clear effects on uncertainty of increasing the pixel size. The first is to reduce the *bias* in the parameter estimate that is induced by positional uncertainty. For β_0 and

β_1 this may lead to the bias occurring in the opposite direction to that observed for the 2 m pixels. However, unless the sample size is very small (10 or fewer measurements per target), the magnitude of this bias is much smaller than it was for the 2 m pixels. The second effect is to reduce the magnitude of the *variation*, as measured by the variance and interquartile range, in the parameter estimates that is induced by positional uncertainty. These effects are most pronounced when moving from the 2 m to 4 m pixels. However, the effect is much less clear when considering the 4, 6 and 8 m pixels, where the bias and magnitude of variation are similar.

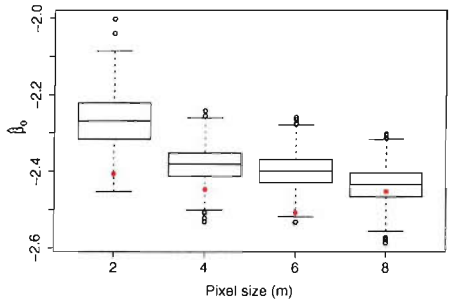
It is obvious that the variation introduced by positional uncertainty arises because field measurements are paired with different pixels. The positional uncertainty that was introduced was distributed $N(0, 4)$. Hence, when the measurement location is perturbed, for the 2 m pixel, there is a high probability that it will be repositioned within an adjacent pixel. However, when a larger pixel size is considered, the probability that the field measurement is repositioned within the same pixel increases. A tentative conclusion is that this is similar for the 4, 6, and 8 m pixels. Hence, if the user wishes to choose the smallest pixel size that is commensurate with the greatest reduction in variation attributed to positional uncertainty, the implication is that they should choose a pixel size on the same scale as the positional uncertainty. In this case, one would choose a pixel size of 4 m. This argument may also be phrased in reverse. If the user wants to maximise accuracy for a given pixel size, will need to adjust the accuracy of their surveying accordingly.

The 4 m pixel is comparable in size to the positional uncertainty that was introduced (*i.e.*, $N(0, 4)$). The rationale for this statement is that a variance of 4 equates to a standard deviation of 2. Since, from the normal distribution we know that approximately 68% of samples lie within 1 sd and 95% lie within 2 sd. Hence, if a point is located at the centre of a 4 m $\times$ 4 m pixel and its location is then perturbed, most perturbations would remain within the pixel. This would tail off for larger pixels. However, if the user requires a 1 m pixel size, then a positional uncertainty of $N(0, 1)$ might be optimal. This assertion needs to be backed up by further experimentation and lies outside the scope of this thesis. This is highlighted as an area for future research.

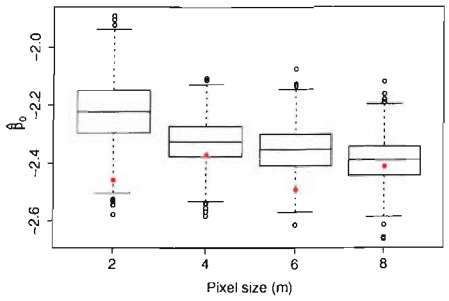
In the absence of positional uncertainty, increasing the pixel size does lead to a change in the estimated parameter values. Hence, as with sample variability, the *position* of the boxplot is affected by the change in pixel size, but the shape is affected by positional uncertainty.

6.2.4 Summary and advice for the practitioner

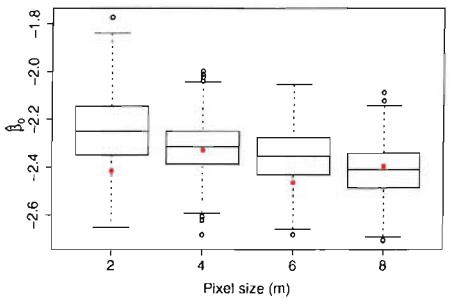
In summarising, it is useful to consider the hypotheses given in the opening section of this chapter. These are discussed below.



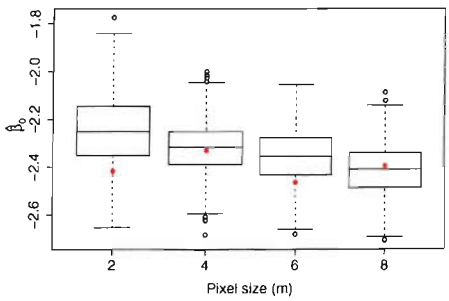
(a) Full data set.



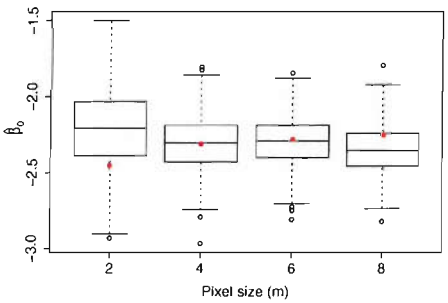
(b) 100 measurements per target.



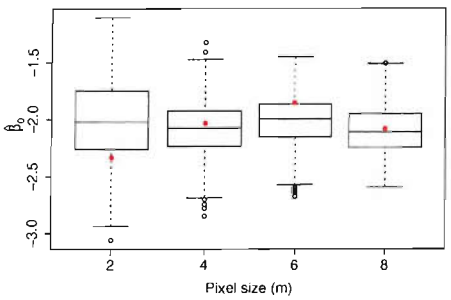
(c) 50 measurements per target.



(d) 30 measurements per target.



(e) 20 measurements per target.

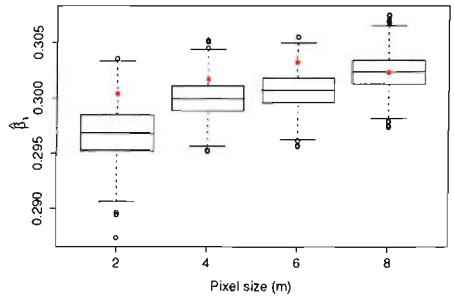


(f) 10 measurements per target.

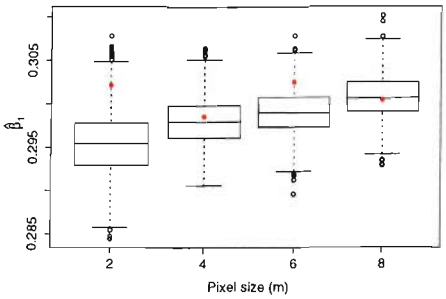
FIGURE 6.5: Box-plots showing variation in $\hat{\beta}_0$ for different sized pixels following the introduction of positional uncertainty, $\varepsilon \sim N(0, 4)$ (point-pixel approach). The red circle indicates the estimated parameter value in the absence of positional uncertainty.

1. It was hypothesised that positional uncertainty would lead to a bias in the estimate of β_0 (positive bias) and β_1 (negative bias), relative to the situation where there was no positional uncertainty. This was found to be the case for the 2 m pixels, except for very small sample sizes (less than 20 measurements per target), where the parameter estimates were highly unstable. When the larger pixels were examined, the bias was reduced considerably and, in some cases was reversed. However, this reversal was small in magnitude.

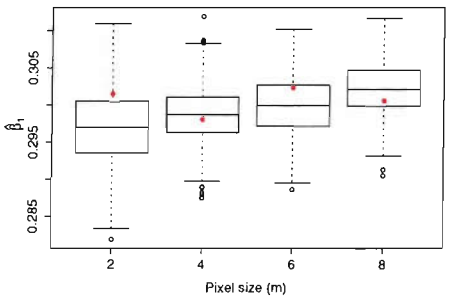
2. It was hypothesised that positional uncertainty would lead to a positive bias in the estimate of σ^2, relative to the situation where there was no positional uncertainty.



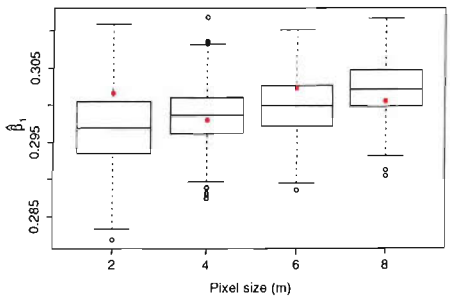
(a) Full data set.



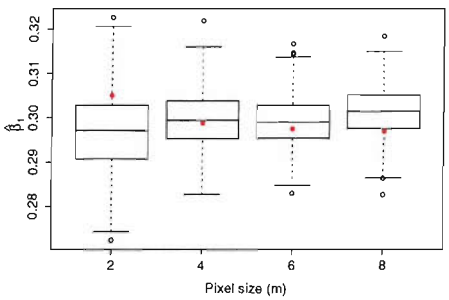
(b) 100 measurements per target.



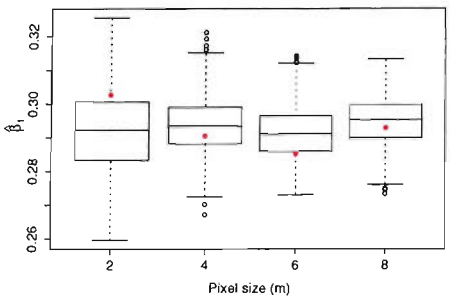
(c) 50 measurements per target.



(d) 30 measurements per target.



(e) 20 measurements per target.



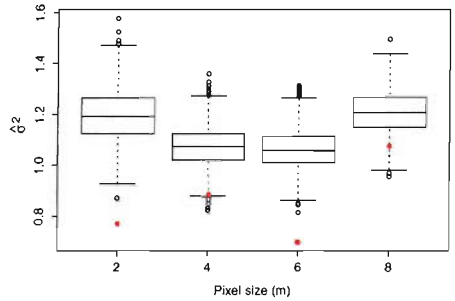
(f) 10 measurements per target.

FIGURE 6.6: Box-plots showing variation in $\hat{\beta}_1$ for different sized pixels following the introduction of positional uncertainty, $\varepsilon \sim N(0, 4)$ (point-pixel approach). The red circle indicates the estimated parameter value in the absence of positional uncertainty.

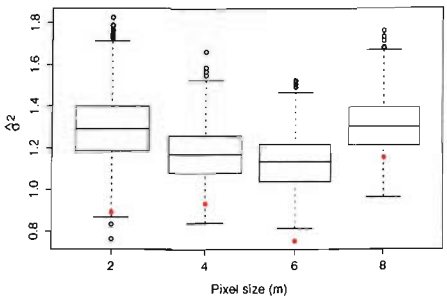
This was found to be the case for all sample sizes and pixel sizes. The implication of this is a reduction in the accuracy in parameter estimation and prediction.

3. It was hypothesised that positional uncertainty would lead to uncertainty in the parameter estimates, as manifested by variability in the parameter estimates between different realisations of positional uncertainty. This was found to be the case for all sample sizes and pixel sizes. The implications are discussed below.

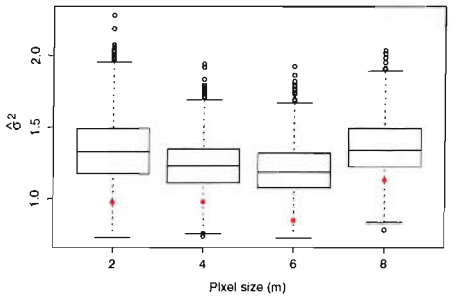
4. It was hypothesised that decreasing the sample size would enhance the previous three effects of positional uncertainty. However, there is no clear pattern to the bias in the estimation of β_0, β_1 and σ^2 with decreasing sample size *that can be attributed*



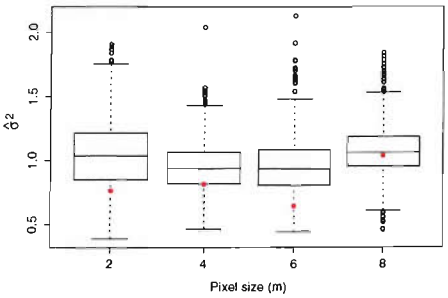
(a) Full data set.



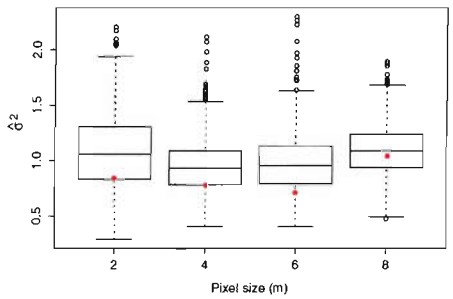
(b) 100 measurements per target.



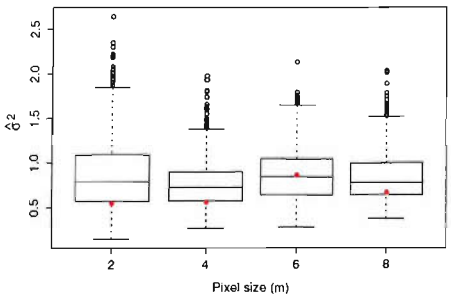
(c) 50 measurements per target.



(d) 30 measurements per target.



(e) 20 measurements per target.



(f) 10 measurements per target.

FIGURE 6.7: Box-plots showing variation in $\hat{\sigma}^2$ for different sized pixels following the introduction of positional uncertainty, $\varepsilon \sim N(0, 4)$ (point-pixel approach). The red circle indicates the estimated parameter value in the absence of positional uncertainty.

to positional uncertainty. For sample sizes less than 20, the estimates were unstable. The variability in the parameter estimates for different realisations of the positional uncertainty (hypothesis 3) clearly increases with decreasing sample size. This is the case for β_0, β_1 and σ^2. The implication is discussed below.

5. It was hypothesised that increasing the pixel size would reduce the effect of positional uncertainty. This was clearly the case when increasing the pixel size from 2 m to 4 m. However, this was less apparent for the transition from the 4 m to 6 m and 8 m pixels.

These results are of consequence, because they lead to variability in the parameter esti-

Parameter		Pixel Size 2 m	4 m	6 m	8 m
β_0	bias ± 0.25	20	5	5	5
	precision ± 0.50	20	10	10	10
β_1	bias ± 0.010	30	5	20	10
	precision ± 0.005	50	30	30	30

TABLE 6.2: Table showing the minimum sample size (per target) required to meet the uncertainty criteria (point-pixel approach).

mates. The value of $\hat{\sigma}^2$ is informative because it allows quantification of uncertainty in estimation and prediction. However, variability in $\hat{\beta}_0$ and $\hat{\beta}_1$ will concern the practitioner, since they affect the position of the regression line and, therefore, the prediction of reflectance. This is illustrated in figure 6.8. This shows the regression line for the full data set and 2 m pixels ($\hat{\beta}_0 = -2.4164$, $\hat{\beta}_1 = 0.3007$) together with the line for $\hat{\beta}_1$ ± 0.01. For a pixel value of 30 (grass) this gives a range from 6.31 to 6.91%, but for a pixel value of 60 (urban surfaces), this gives a range from 15.01 to 16.23%.

Moving on from this, it is possible to establish uncertainty criteria that can be used to establish the minimum sample size for a given pixel size, given the expected positional uncertainty. For β_1 one might set a *bias* of ≤ 0.005 and a precision of $mean(\hat{\beta}_1) \pm 0.01$. Hence, the mean and standard deviation ($sd(\cdot)$) of the 1000 simulations is calculated. The $mean(\hat{\beta}_1) \pm 2 \times sd(\hat{\beta}_1)$ was used to establish the precision. The bias was established by computing difference between the mean and $\hat{\beta}_1$ as calculated in the absence of positional uncertainty. A similar procedure was used for β_0, with the bias was set at 0.25 and the precision at ± 0.5. This process is illustrated in table 6.2. From this table it can be established that, according to these uncertainty criteria, 50 measurements per target are required for the 2 m pixels and 30 measurements per target for the 4 m pixels. No advantage is obtained by going for a 6 m or 8 m pixel. Indeed, this may confuse the issue because the difference in support size between the pixel and field measurements is very large.

The above discussion has focused on the estimation of β_0 and β_1. However, to quantify the accuracy of prediction it is necessary to consider the uncertainty in the estimate of σ^2. The effect of this was illustrated in section 6.2.1. For the 2 m pixels a sample size of $n50$ per target results in a bias of 0.4 and an imprecision[2] of 0.9. However, increasing the sample size to $n100$ reduces the imprecision to 0.6, whilst increasing it to $nFull$ further reduces the imprecision to 0.4. For the 4 m pixels, $n30$ gives an imprecision of 0.8, whilst increasing the sample size gives an imprecision of 0.7, 0.5 and 0.3 for $n50$, $n100$ and $nFull$ respectively. Because of the importance of accurate estimation of σ^2, a minimum sample size of 50 or preferably 100 measurements per target is recommended.

[2]Note that, in this case, imprecision refers to $mean(\hat{\sigma}^2) \pm 2 \times sd(\hat{\sigma}^2)$ whereas, in statistics the precision is taken to be the inverse of the variance.

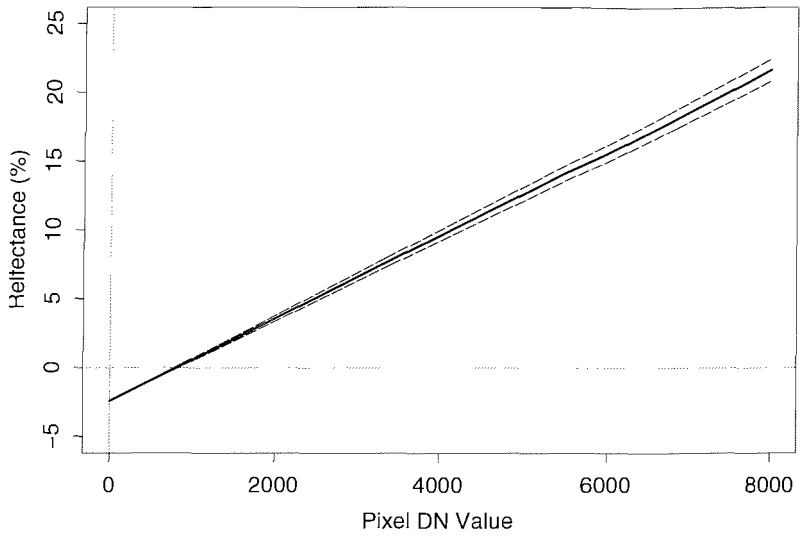


FIGURE 6.8: Shows the estimated regression line for the full data set and 2 m pixels (thick line). The dashed lines show the regression line for $\hat{\beta}_1 \pm 0.01$.

6.2.5 Conclusions

The analysis discussed in this section has covered the potential impact of uncertainty in the location of field measurements on the outcome of the ELM when implemented using the point-pixel approach. It has been demonstrated that introducing positional uncertainty introduces uncertainty into the parameter estimates in the form of both bias and imprecision. This is exacerbated by reducing the sample size, but can be ameliorated by increasing the pixel size. Overall, the impact on the estimates of β_0 and β_1 were small for large sample sizes (50 or larger) although they could be substantial for smaller sample sizes. However, the bias in the estimate of σ^2 was substantial for all sample sizes and this leads to widening of the confidence and prediction intervals on the regression line. The only way to reduce this bias was to increase the pixel size.

Overall, this analysis emphasises the importance of accurately recording the location of field measurements that are to be paired with remotely sensed data. If that is not feasible, then a large sample is required to produce accurate estimates of β_0 and β_1 and the user will have to accept the large values for σ^2. It is recommended that at least 50 and preferably 100 measurements per target are taken. This coincides with the recommendation for the point-pixel approach in the absence of positional uncertainty. *However*, it should be recognised that the value of σ^2 is larger than it could be. Hence, this recommendation should *not* be taken as a reason not to implement an accurate surveying strategy.

Finally, it should be noted that the degree of perturbation for the analysis presented in this section was large ($N(0, 4)$). In reality it would be hoped that positioning could be carried out more accurately than in figure 6.1.

6.3 Block-pixel

This section considers the implication of positional uncertainty when the ELM is implemented using the block-pixel approach. In this section, the aggregation was performed using block kriging. In addition to evaluating the implication for the ELM, this section is interesting since *"en route"* it evaluates the impact of positional uncertainty in a geostatistical context. This has received only limited attention in the literature (Gabrosek & Cressie 2002).

The analysis discussed here was presented at the Fourth International Conference on Geostatistics for Environmental Applications (geoENV IV). It was subsequently published (Hamm et al. 2004) in an edited collection (Sanchez-Vila et al. 2004). It actually represents an earlier stage than much of the other analysis presented in this thesis. It uses a different airborne data set to that used elsewhere in the thesis and adopts a slightly different methodology. In particular, the Bayesian, rather than maximum likelihood (ML) approach to model-based geostatistics (MBG) was adopted. Hence, in one sense, this section stands apart from the rest of the thesis and is rather less refined. However, the analysis and results are highly relevant in the context of the rest of the thesis and it is presented as published.

This analysis was conducted before the ATM data were made available. Hence, unlike other analysis presented in this thesis, it was undertaken using CASI data. Hence, MMR band 1 (approximately 420 nm – 530 nm) is paired with CASI band 1 (450 nm with a width of 20 nm). This pairing is not ideal since the match between the spectral bandpasses is less accurate than it is for the ATM data. However, it is sufficient for exploring the impact of positional uncertainty on the outcome of the ELM.

The process of aggregation using block kriging was discussed for the individual targets in section 5.4.2. The analysis conducted in this section differed for the concrete target in that the concrete target was not split in two. Rather, it was treated as a single target and several outliers were omitted, yielding a histogram that approximated a normal distribution. The sample variogram of the single concrete surface is given in figure 6.9. The sample variograms for the asphalt and grass GTs are as presented in section 5.4.2.

As in section 6.2, perturbed data sets were created using the approach outlined in section 6.1. However, analysis was performed for two sets of perturbations, $\varepsilon \sim N(0, 0.25)$ and $\varepsilon \sim N(0, 4)$. Hence, two sets of 1000 perturbations were produced. However, unlike the previous section the analysis was *not* conducted for a reduced data set. This is because a large data set is required for geostatistical analysis.

The variogram modelling and block kriging presented in section 5.4.2 demonstrated the utility of both classical geostatistics and MBG. Although the two approaches might lead to the utilisation of different models, the cross-validation showed the accuracy of both approaches to be similar. In this section the MBG approach was adopted exclusively.

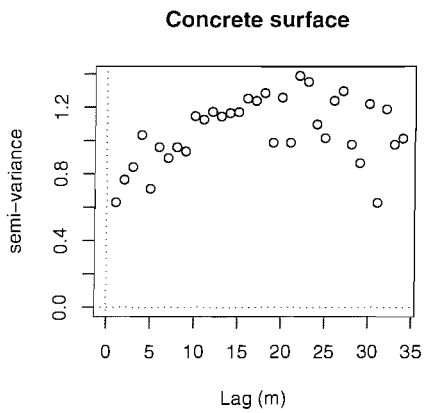


FIGURE 6.9: Omnidirectional sample variogram for the concrete GT (MMR band 1). Outliers have been omitted so that the histogram approximates a normal distribution.

The MBG approach estimates the variogram parameters without having to first compute a sample variogram (Diggle et al. 1998). The analysis conducted required that the variogram be modelled for *each* of the 1000 perturbations. Hence an automated approach was required and this could be performed more straightforwardly using MBG.

In this section, the *Bayesian* implementation of MBG was utilised rather than the ML implementation (see section 2.2.2.5). Given the absence of prior information, estimation and prediction were undertaken with non-informative priors for the variogram parameters (Diggle & Ribeiro Jr. 2002, Diggle et al. 2003). An exponential model with a first order trend on location was adopted. The analysis presented in section 5.4.2 does not always return the exponential model as the most accurate model. However, it is not substantially sub-optimal in any of those cases and its smooth properties aided computation. The spherical model is actually comprised of two model – one up to, and the other after, the range. The non-differentiability at $h = \phi$ can lead to computational problems with maximisation (Diggle et al. 2003). This was led to computational failure when automating the variogram modelling.

The Bayesian MBG implementation was not pursued after this section for various reasons. First, it was found to give comparable results to the ML approach. Second, it is both conceptually and computationally less straightforward to implement than MBG-ML. Third, the application does not exploit the full power of the Bayesian implementation and hence adopting a simpler approach was deemed sufficient. Examples of the full power of the Bayesian implementation come when prior information is available or the objective is to predict statistical properties of a surface (such as exceedence probabilities) (Diggle et al. 1998, Diggle et al. 2003). Furthermore, the Bayesian approach offers the flexibility to develop hierarchical models (Diggle et al. 2003, Wikle 2003), a topic that is given some consideration in chapter 8.

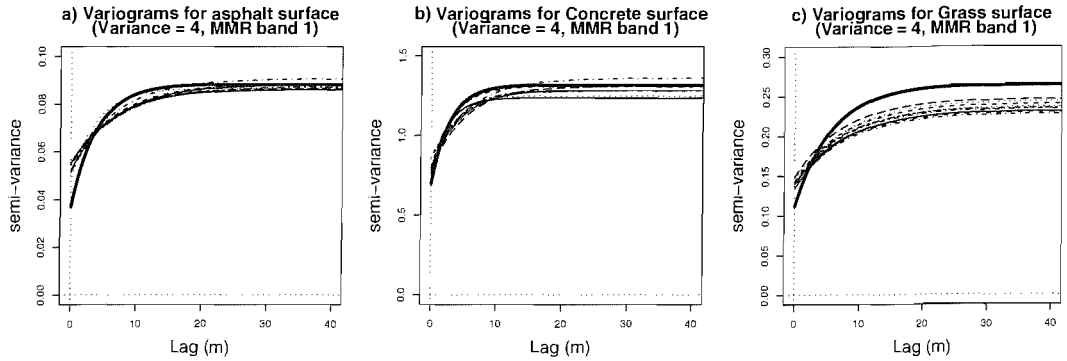


FIGURE 6.10: Modelled variograms for the three surfaces in the presence of positional uncertainty (a) asphalt, (b) concrete and (c) grass. The thick line is for the "original" (i.e. unperturbed) data set and the remaining lines are each for different perturbed data sets (in this case $\varepsilon \sim N(0, 4)$).

6.3.1 The effect of positional uncertainty on the variogram

The modelled variograms for the original (i.e. unperturbed) data set and for the perturbed ($\varepsilon \sim N(0, 4)$) data set are shown in figure 6.10. From theory, it is expected that the sill of the variogram will not change in the presence of positional uncertainty. However, the range may increase and the variogram will increase at short lags (Atkinson 1996, Gabrosek & Cressie 2002) (see also section 2.3). This theoretical argument supports intuition, which would suggest less clear evidence for spatial structure in the presence of positional uncertainty. The expected effects are borne out for the asphalt and concrete surfaces, but not for the grass surface, where there is a decrease in the sill. Although unexpected, the latter result is not inconsistent with other *experimental* results (Atkinson 1996). The reason for this result is unclear, although it may be noted from section 5.4.2.3 that the grass surface showed evidence of anisotropy. This was *not* taken into account in the variogram modelling conducted for the analysis presented in this section.

It is expected that the increase in the variogram at short lags will lead to decreased accuracy in point-kriged predictions. This will be exacerbated since the data points that are used for kriging will be incorrectly located. The effect on block-kriged predictions is likely to be less significant. However, this impact cannot be evaluated in an absolute sense, since there are no measurements of reflectance available for the blocks. In the context of this investigation, it is the eventual impact of the positional uncertainty on the output of the ELM that is important, *not* the absolute effect on the block-kriged surface. Finally, it is unclear what the impact of these results would be on the outcome of conditional simulation.

6.3.2 The effect of positional uncertainty on the ELM regression

The regression was performed using the co-located kriged blocks and pixels. The estimated parameters for the original (unperturbed data set) are given in table 6.3, together with

	$\varepsilon = 0$	$\varepsilon \sim N(0, 0.25)$			$\varepsilon \sim N(0, 4)$		
	Orig	P.1	P.2	P.3	P.1	P.2	P.3
β_0	-0.76	-0.76	-0.75	-0.74	-0.75	-0.68	-0.71
β_1	0.29	0.29	0.29	0.29	0.29	0.29	0.29
σ	0.53	0.53	0.53	0.53	0.53	0.53	0.52
R^2	0.98	0.98	0.98	0.98	0.98	0.98	0.98

TABLE 6.3: Estimated parameters for the ELM, implemented by pairing blocks and pixels. Orig. indicates the results for the unperturbed data sets. P.1, P.2 and P.3 indicate different realisations of the perturbed surface for $\varepsilon \sim N(0, 0.25)$ and $\varepsilon \sim N(0, 4)$.

examples of estimated parameters for the perturbed data sets (for both $\varepsilon \sim N(0, 0.25)$ and $\varepsilon \sim N(0, 4)$). The examples given are for illustration and are consistent with other results.

The results presented in table 6.3 suggest that the introduction of positional uncertainty has a negligible effect on the estimate of β_1 and σ. However, where ε is large ($\varepsilon \sim N(0, 4)$), this may have a small effect on the estimate of β_0. For many remote sensing applications this change in β_0 may be unimportant, since it introduces a small bias of less than 0.1% (for 95% of cases)(Smith & Milton 1999). Overall the lack of sensitivity of the parameter estimates to substantial positional uncertainty is encouraging.

It should be realised that the blocks are derived from a kriged surface. Using blocks that are derived from conditionally simulated surfaces are likely to lead to a larger and more realistic estimate of the variance in the regression model (see section 5.4.2). This remains as a potential avenue for further research.

6.3.3 Summary and advice for the practitioner

The lack of sensitivity of the parameter estimates to substantial positional uncertainty was unexpected. The implication of this is that the user can opt for a cheap and fast mechanism for recording location and that this will not substantially affect parameter estimation for the ELM. However, it must be recognised that the geostatistical approach to block prediction has already been shown to be problematic, as follows:

1. The results of the geostatistical analysis were disappointing for all three surfaces. This was because of the low accuracy of prediction, as evidenced by the cross-validation diagnostics, despite the clear evidence for spatial structure. The estimates of the ELM parameters were, hence, primarily determined by the *between-target* variability rather than the *within-target* variability. By decreasing the accuracy of the geostatistical implementation an inaccurate prediction has merely been made slightly less accurate. However, the parameter estimation is still governed by the between-target variability. If there was a more accurate mechanism for predicting the reflectance of blocks, the effect might be more substantial.

2. In section 5.4.2 substantial reservations were raised regarding the use of kriged, or indeed conditionally-simulated, blocks for the ELM and empirical regression models more generally.

6.4 Summary and conclusions

This chapter evaluated the impact of positional uncertainty on parameter estimation and prediction for the ELM. The *FP model* of positional uncertainty was adopted whereby the *attribute* is correct, but the *location* is recorded with error (Gabrosek & Cressie 2002) (see also section 2.3). The rationale for choosing the FP model over the CP model was discussed in section 6.1. However, whichever framework is considered appropriate it is argued that the emergent effect will be the same – although an assessment of this could be the subject of future investigations using simulated data or densely sampled field data. For the point-pixel approach it leads to the possibility of a measurement being paired with the wrong pixel. For the geostatistical block-pixel approach it will be manifested as measurement error and lead to an increase in the variogram at short lags (Atkinson 1996) with implications for the accuracy of kriged predictions. The distinction between the CP and FP models becomes important when it is corrected using a specific model (e.g., Gabrosek & Cressie 2002) or when the interest lies in the location of specific items, such as the vertices of polygons (Brown & Heuvelink 2007).

Overall, the analysis demonstrated that, for substantial positional uncertainty ($\varepsilon \sim N(0, 4)$, the point-pixel approach is sensitive to positional uncertainty at *all* sample sizes. For sample sizes larger than 50 this is most clearly manifested by an increase in the estimate of σ^2, with consequent impacts on the confidence intervals. In order to minimise the effects of bias and imprecision in the estimates of the regression parameters it was recommended that at least 50 and preferably 100 measurements should be taken per target. This is comparable to the number recommended in the absence of positional uncertainty. However, importantly, this *does not* incorporate any sample variability. Furthermore, the user will still have to tolerate an increase in uncertainty associated with the increase in σ^2.

For the geostatistical block-pixel approach the modelled variograms for asphalt and concrete supported the changes that are expected from theory and intuition. However, the changes in the variogram do not lead to sufficient changes in the kriged surfaces to be manifested in changes to the estimated ELM parameters. This result applies for both a low ($N(0, 0.25)$) and high ($N(0, 4)$) level of positional uncertainty. This is an encouraging result. However, it needs to be set in the context of wider concerns about the use of geostatistical block-kriging and needs further examination.

Finally, it should be noted that the point-pixel and block-pixel approaches have different manifestations in the ELM. When the point-pixel approach is implemented positional uncertainty leads to a change in the x-value associated with a given y value. For the block-pixel approach any change in the kriged blocks will lead to a change in y.

Overall, it should be recognised that the impact of positional uncertainty on geostatistical analysis and empirical regression models for remote sensing is a complex topic that has received limited attention in the literature. In the context of the research conducted for this thesis the following areas are worthy of investigation and provide possible avenues for future research:

1. For the point-pixel approach the sample size was held fixed and the location perturbed. It would be interesting to explore the interaction between positional uncertainty and sample variability.

2. For the point-pixel approach only a large perturbation was investigated ($\varepsilon \sim N(0, 4)$). It would be interesting to investigate the impact of both less and more extreme positional uncertainty. Furthermore, this would allow examination of the hypothesis that the impact of positional uncertainty does not decrease for pixels that are larger (in size) than the scale of positional uncertainty.

3. From a geostatistical perspective, this area would benefit from a more complete analysis of the impact of positional uncertainty for prediction and simulation. To date, most research has focused on the impact on the variogram (Atkinson 1996, Gabrosek & Cressie 2002) with the impacts on kriging being assessed using simulated or well understood data sets (Gabrosek & Cressie 2002, Cressie & Kornak 2003). The impact on prediction and simulation was not, in itself, addressed in this chapter, since interest focused on the impact on the ELM.

4. Chapter 5 recommended the simple averaging approach to aggregation from points to blocks. However, the effect of positional uncertainty on that approach was not investigated. It is expected that the results will be less extreme than those found for the point-pixel approach.

Chapter 7

Analysis III: modelling issues

In chapter 5, two alternative methods for implementing the ELM were proposed. These were the point-pixel and block–pixel approaches. There were two problems with the implementation of the regression for the ELM which, although highlighted, have not yet been dealt with. These are:

1. The residuals from the regression model were clearly heteroskedastic – that is to say the magnitude of the residuals clearly varied among GTs.
2. It was demonstrated that there was spatial structure across the GTs for both the MMR reflectance data and ATM radiance data (section 4.5.1 and section 4.5.2). Hence, it is likely that the residuals from the regression model are also autocorrelated.

Both of these issues present problems because they contravene fundamental assumptions of the regression model – that the residuals are *independent* and *identically distributed.* Hence, they need to be addressed and their impact evaluated. Addressing these issues should be seen in a much broader context than the ELM. Regression models are very widely used in remote sensing and GIS. Yet the above issues are rarely dealt with. An oft–cited alternative to regression is geostatistics. However, as discussed in section 2.4, this, in itself, may not be a useful alternative.

In addition to theoretical concerns about the validity of the regression model, the above two issues are important since, if they are not addressed, they may lead to (i) inaccurate estimators of the regression coefficients, (ii) inaccurate estimation of the uncertainty in estimation and (iii) inaccurate estimation of the uncertainty in prediction. This is discussed in section 2.4.

The regression plots presented in chapter 5 illustrated uncertainty in estimation through confidence intervals and uncertainty in prediction through prediction intervals. Derivation of these was discussed in chapter 2.1. Examples of estimation and prediction uncertainty

focused on a single *pixel.* It did not show how this might be extended to images. Predicting over an image is more complex, because realisations cannot be drawn from individual pixels because they are not independent. Hence, although it is relatively straightforward to predict the expected or *mean* image, it is not clear what individual realisations of the random function or, for example, the 95% prediction bands should look like.

This chapter begins by recapping the issue of heteroskedasticity and discusses means for addressing this. It then moves on to applying and testing these in the context of the ELM. The results of that analysis are then evaluated. Section 7.4 considers mechanisms for accounting for spatial autocorrelation in the residuals. These were introduced in section 2.2.2.4 and section 2.4. This section works with simulated data to develop and test the mechanism. Section 7.5 follows on from this, but extends the model to cope with heteroskedasticity. The model was developed and tested with simulated data, since this allows a controlled environment to test the accuracy of the model. Although such models have appeared in the literature, there has been only limited evaluation on simulated data (e.g., Lark 2000). Similarly *no* examples of it being applied to heteroskedastic data have been found in the remote sensing and GIS literature, although there are some early examples of it being applied in epidemiology (Cook & Pocock 1983). Hence, this analysis on simulated data is valuable in its own right. Finally, the combined heteroskedastic-correlated error model is applied to the ELM (section 7.6.1).

7.1 Non heteroskedastic residuals

The case where the regression residuals are not identically distributed is called *heteroskedasticity.* In extreme cases this is manifested in the regression plots (e.g., figure 5.7). However, it is better illustrated using standard S-Plus regression diagnostic plots, as presented in figure 7.1. This shows the residuals plotted against the fitted (predicted) values (top left), the square-root of the standardised residuals against the fitted values and the QQ plot (top right). If the residuals are identically distributed they should show a similar distribution around zero for all fitted values. If they are normally distributed the QQ plot should approximate to a straight line. The standardised residuals are standardised by the standard-deviation (bottom left). Hence, if weighted least squares (WLS) is successfully applied (as discussed below), this plot will show constant variance, even if the residuals *vs* fitted values plot does not. Finally, the Cook's distance plot (bottom right) gives an indication of whether specific data points exert undue influence on the regression and might be considered outliers. Figure 7.1 gives clear evidence of heteroskedasticity and is discussed further below.

A common approach for dealing with heteroskedastic residuals is to adopt the approach of "weighted linear regression" which may be performed using weighted least squares (WLS) (Sen & Srivastava 1990, Neter et al. 1996, Gelman et al. 2004). In such cases, the

variance may be described as:

$$\sigma_i^2 = \frac{1}{w_i}\sigma^2 \tag{7.1}$$

where the i denotes a particular observation and σ_i^2 is the variance of the error term associated with that observation. The w_i are weights attached to each observation and σ^2 is a constant term. The "weights" are so called because observations with a high variance are "downweighted" in the parameter estimation (Sen & Srivastava 1990). If the relative magnitudes of the σ_i^2 are known, these can be used directly and σ^2 will approach 1. Hence, σ^2 acts as a constant of proportionality. If the w_i are not known, they need to be estimated.

If the w_i are known (or can be estimated) they can be used to populate a $n \times n$ (n is the number of data points) diagonal weights matrix, $\mathbf{W}$, as follows: This matrix is then used to "weight" the estimation of the regression coefficients $\beta = (\beta_0, \beta_1)^T$), as follows:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y} \tag{7.2}$$

The mean square error (MSE_w) is then estimated as:

$$MSE_w = \frac{\sum w_i(y_i - \hat{y}_i)^2}{n-2} \tag{7.3}$$

where MSE_w is an estimator of σ^2.

The confidence interval for a new values of x, $\mathbf{x_0} = (1, x_0)^T$ is given as (Sen & Srivastava 1990):

$$\hat{y}_0 \pm t_{1-a/2}\hat{\sigma}^2[\mathbf{x_0}^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{x_0}]^{1/2} \tag{7.4}$$

where $1-a$ is the desired confidence level (e.g., 95%) and the prediction interval as:

$$\hat{y}_0 \pm t_{1-a/2}\hat{\sigma}^2[\frac{1}{w_0} + \mathbf{x_0}^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{x_0}]^{1/2} \tag{7.5}$$

Neter et al. (1996) provide various suggestions for predicting the w_i by modelling the residuals, e_i, as a function of predictor variables x_i. This is based on the fact that e_i^2 is an estimator of σ_i^2 (and $|e_i|$ and estimator of σ_i). Hence, we set $e_i^2 = \hat{\sigma}_i^2$ and $|e_i| = \hat{\sigma}_i$. For example, one of the following strategies might be adopted:

$$\hat{\sigma}_i = \alpha_0 + \alpha_1 x_i \tag{7.6}$$

or

$$\hat{\sigma}_i^2 = \alpha_0 + \alpha_1 x_i \tag{7.7}$$

The weights are then given as the inverse of σ_i^2, i.e., $w_i = 1/\sigma_i^2$ and σ^2 is modelled as the MSE. It is expected that $MSE_w \approx 1$. The x_i above could be replaced by another variable, including $\hat{y}_i$. In order to obtain the e_i, one might begin with the residuals obtained from OLS and then iterate to obtain a new set of residuals.

Neter et al. (1996) also note that, where replicate measurements are made at different levels of the predictor variable, these replicates can be used to estimate the weights. If a sufficiently large number of replicates have been obtained, then the weights can be estimated directly. Alternatively, the sample variances or standard deviations can be regressed against the appropriate predictor variable, as described above. For experiments, replicates can be included in the experimental design although for observational studies, near replicates may be used. A similar approach is suggested by Gelman et al. (2004), who advocate dividing the data set into batches of similar variance. In the situation reviewed in this study, the batches might correspond to different land cover types.

Finally, it could be assumed that the weights are directly related to a third variable, and treated as if they are known (Sen & Srivastava 1990, Neter et al. 1996).

For the problem illustrated with the ELM the residuals do not tend to vary as a function of x (*i.e.*, radiance). However, the magnitude of the residuals appears to be related to the variability of the target in question. On this basis, three solutions for estimating the weights are proposed:

1. Split the data up into batches determined by the target cover type, following the procedure of Gelman et al. (2004).
2. Model the weights as a continuous function, similar to the procedures suggested in equations 7.6 and 7.7, but substituting another measure for x. Appropriate measures might be the standard deviation or coefficient of variation of the pixels surrounding the pixel of interest. For example, a 3×3 window might be defined, where the pixel of interest is the centre pixel.
3. Some hybrid of the above two.

The first approach is appropriate for parameter estimation. However, it is important to note that it is only of limited use for prediction. This is because it is not possible to model how σ_i^2 will look for targets that were not included in the model building data set. Hence, for prediction it is necessary to know w_0 for a "new" x–value where prediction is required (see equation 7.5). Because the weights are estimated using the data, rather than being known *a priori* or estimated separately this remains unknown. In order to calculate the prediction variance it is necessary to assume a value for w_0 or estimate it using some method (such as the second approach). The second approach requires only the measure determined by the surrounding pixels, hence it is possible to estimate the weight at any point in the image. It can, therefore, be used for prediction. Nevertheless, the first approach is appropriate where the primary objective is to accurately estimate the regression line and associated confidence limits. Furthermore, this provides a useful benchmark against which other methods can be evaluated.

Following from the above discussion, three approaches for estimating the weights were considered.

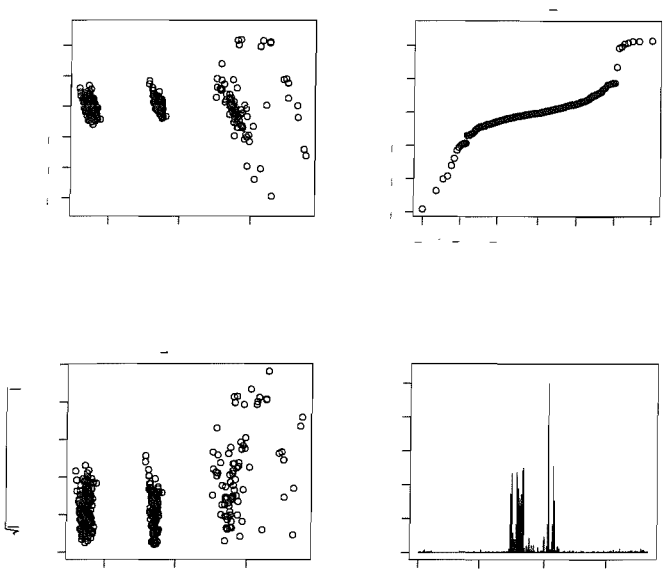


FIGURE 7.1: Diagnostic plots of the residuals for MMR band 1 (block–kriged prediction) regressed on ATM band 2.

1. The weights were estimated using the variance of the OLS residuals for each target. As described above, this approach is useful for estimation but not for prediction.

2. The residuals were modelled as a function of the variance in a 3×3 window surrounding the pixel of interest.

3. The residuals were modelled as a function of the variance in a 3×3 window surrounding the pixel of interest. They were then averaged for individual GTs and cover types.

When the simple averaging approach to aggregation from field measurements to pixel sized blocks is adopted (figures 7.2 and 7.3) the same conclusions are reached. For figure 7.2 all pixels containing *one or more* field measurements are included, whereas all pixels containing *two or more* field measurements are included for figure 7.3. This was discussed in section 5.4.1. For the analysis conducted in this chapter, the former data set is used, since it provides a larger sample size. However, the implications of this are discussed and given further consideration in section 8.1.

Subsequent analysis presented is for MMR band 1 regressed on ATM band 2.

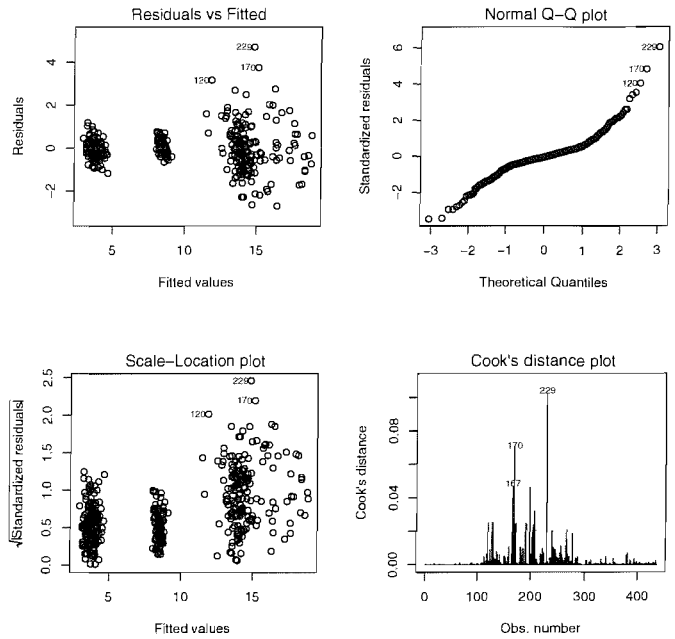


FIGURE 7.2: Diagnostic plots of the residuals for MMR band 1 (simple averaging approach) regressed on ATM band 2. Under this implementation of the simple averaging approach all pixels containing *one or more* field measurements are included.

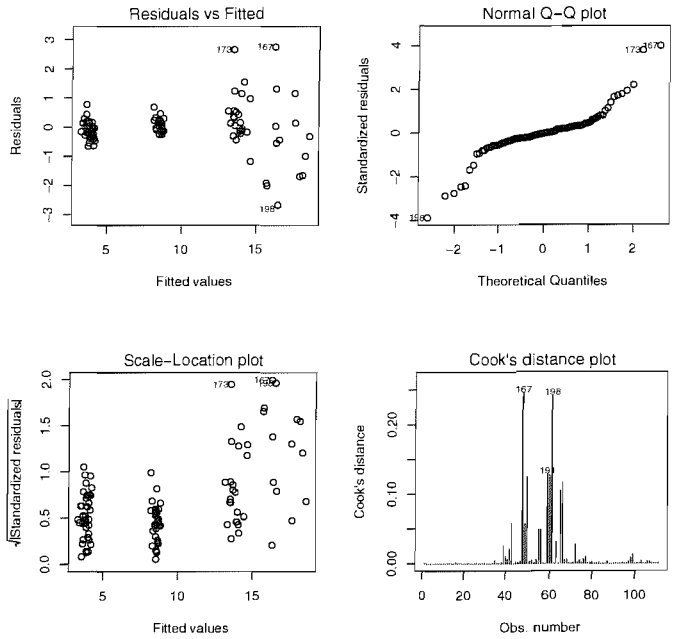


FIGURE 7.3: Diagnostic plots of the residuals for MMR band 1 (simple averaging approach) regressed on ATM band 2. Under this implementation of the simple averaging approach all pixels containing *two or more* field measurements are included.

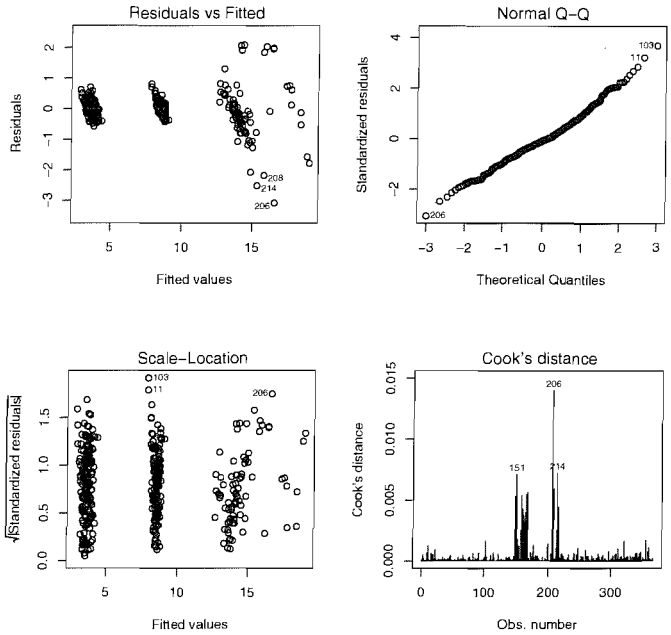


FIGURE 7.4: Diagnostic plots of the residuals for MMR band 1 (block-kriged prediction) regressed on ATM band 2. The regression is implemented using WLS, where the weights are given as the variance of the OLS residuals for each of three cover types (asphalt, concrete and grass).

7.2 Calculating the weights: A strategy for estimation

This section implements the first weighting strategy that was listed above. This follows the approach of Gelman et al. (2004) and splits the data into appropriate "batches". For the ELM the batches correspond to GTs. To calculate the per-GT weight, the variance of the OLS residuals is calculated on a per-GT basis. The reciprocal of the per-GT residual variance gives the weight for that GT.

7.2.1 Kriged blocks

Kriging and then conditional simulation were used to aggregate from points to pixel-sized support. The OLS regression for the kriged blocks is given in figure 7.1. As previously discussed, this shows clear evidence for heteroskedasticity. The OLS residuals were then used to calculate the weights and weighted least squares (WLS) was applied. The diagnostic plots are shown in figure 7.4 and the estimated parameter values are shown in table 7.1. It can be seen the standardised residuals are now approximately normally distributed and homoskedastic. The estimated values of σ_i^2 for each target are also shown and can be seen to differ substantially from the estimated value of σ^2 for OLS. Note that the *relative* weights ($w_i/max(w)$) are also provided, since these allow straightforward comparison between different methods for specifying the weights.

The confidence and prediction bands for $\mathbf{x_0} = (20, 40, 60)^T$ are shown in tables 7.2 and 7.3

Parameter	Estimate	Std. Error	p value
β_0	-2.4634	0.0488	< 0.01
β_1	0.3033	0.0016	< 0.01
σ^2	1.0		
R^2	0.99		
	Asphalt	Concrete	Grass
w_i	20.38	1.01	16.47
$w_i/max(w)$	1.0	0.0494	0.8079
$\sigma_i^2 = \sigma^2/w_i$	0.0500	1.0130	0.0620
σ_i	0.2237	1.0065	0.2489

TABLE 7.1: Table showing the results of the parameter estimation for the ELM regression model after the reflectance measurements have been aggregated using block kriging. The estimates are for MMR band 1 regressed on ATM band 2.

x_o	$\hat{y}_0$	Lower CI	Upper CI	Lower PI	Upper PI
20	3.62	3.58	3.66	2.62	4.62
40	9.63	9.60	9.66	8.63	10.63
60	15.64	15.58	15.70	14.64	16.64

TABLE 7.2: Table showing the predicted value of y_0 and the lower and upper 95% confidence (CI) and prediction (PI) intervals. The values of x_0 correspond to grass, asphalt and concrete respectively for MMR band 1 (aggregated using block kriging) regressed on ATM band 2 using OLS.

for OLS and WLS respectively. This corresponds to grass, asphalt and concrete respectively. The confidence interval quantifies the uncertainty in the position of the regression line and the prediction interval the uncertainty in making a new prediction. By properly accounting for the heteroskedasticity these can be estimated more accurately. Under WLS the confidence intervals are narrower. Hence there is less uncertainty in the position of the regression line. However, it may be noted that, in absolute terms, the confidence intervals are narrow in both cases. As is evident from equation 7.5 the prediction intervals vary from GT to GT as a function of the weight. For asphalt, the interval is narrower, whereas for concrete it is wider than for the OLS. Hence, if heteroskedasticity is not accounted for the level of uncertainty in a predicted result may be over- or under-estimated. It should be emphasised that the prediction intervals cannot be calculated for other surfaces, since the weights are unknown. This is addressed in section 7.3.

x_o	$\hat{y}_0$	Lower CI	Upper CI	Lower PI	Upper PI
20	3.59	3.56	3.64	3.11	4.09
40	9.68	9.64	9.72	9.24	10.12
60	15.76	15.66	15.86	13.78	17.75

TABLE 7.3: Table showing the predicted value of y_0 and the lower and upper 95% confidence (CI) and prediction (PI) intervals. The values of x_0 correspond to grass, asphalt and concrete respectively for MMR band 1 (aggregated using block kriging) regressed on ATM band 2 using WLS.

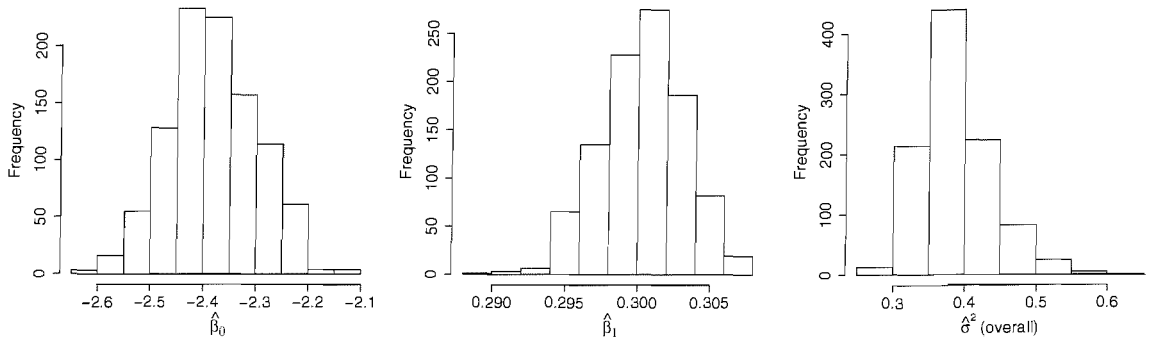


FIGURE 7.5: Histograms showing the estimates of the OLS model parameters for 1000 conditionally simulated realisations of the targets.

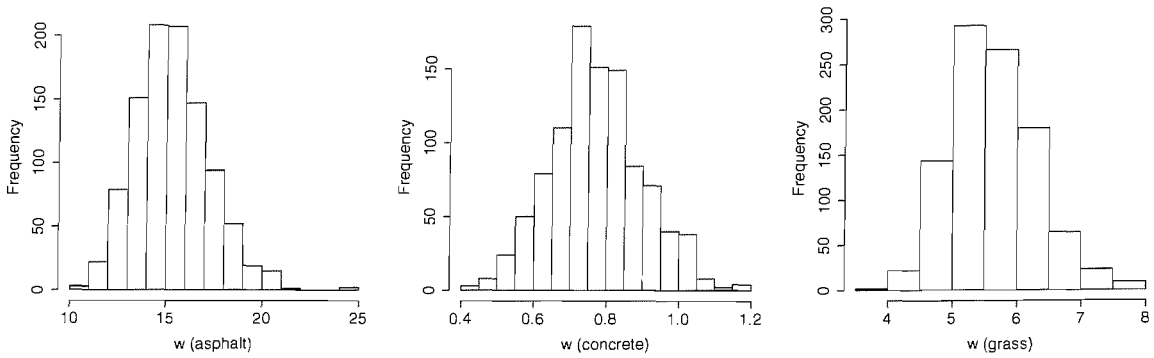


FIGURE 7.6: Histograms showing the estimated weights for 1000 conditionally simulated surfaces where $w_i = 1/Var(e_i)$, where e_i refers the OLS residuals for a specific target, i (*i.e.*, asphalt, concrete or grass).

The results presented above were for the kriged target reflectance. The kriged surfaces show lower variance than the true surface (or possible reality). Therefore, the y value used in the regression will display lower variability than would be expected. This will lead to an underestimate for σ^2 and, possibly, a flatter regression line. However, the magnitude of this latter effect will depend on relative influence of *within* and *between* target variability.

As discussed in section 2.2.2.6 and section 5.4.2, conditional simulation can be used to predict a possible reality. Hence, conditional simulation was used to create 1000 realisations of reflectance for each target. These were then paired with the image data and used in the regression model. The results are represented as histograms. Figure 7.5 shows histograms of the estimates of the parameters, as estimated using OLS. The variance of the residuals for each of the targets was then calculated and the inverse of this value taken as the weight for each target. Histograms of the weights are shown in figure 7.6 and the relative weights are shown in figure 7.7. The relative weights are given as $w_i/max(w)$ and are useful because they allow direct comparison between different methods of estimating the weights. Note that if the relative weights are used in WLS, the resulting inference will be the same, although the estimated value of σ^2 will differ.

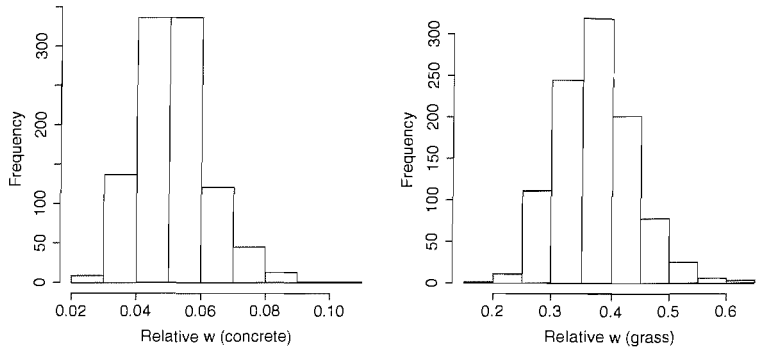


FIGURE 7.7: Relative weights for 1000 conditional simulated targets where $w_i(rel) = w_i/max(w)$. Note that the relative weight for asphalt surface is always 1.

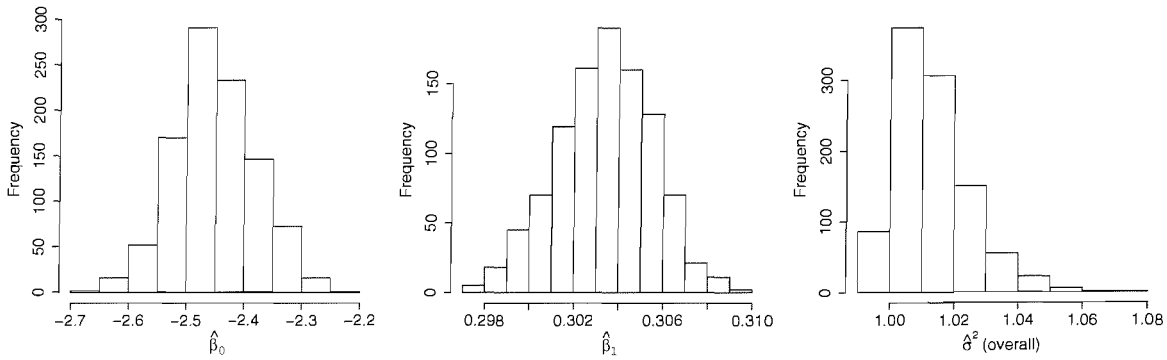


FIGURE 7.8: Histograms showing the estimates of the WLS model parameters for 1000 conditionally simulated realisations of the targets. The weights used were shown in figure 7.6.

For the conditionally simulated surfaces, there is a clear decrease in the weights by comparison to the kriged surface. This reflects the increase in the variance associated with a conditionally simulated surface. However, the story for the relative weights is different. The asphalt surface was always assigned the greatest weight (it has residuals with the lowest variance), hence it has a relative weight of 1. The relative weight for the concrete surface is approximately normally distributed, centred on 0.05 – the same relative weight as that obtained for the kriged surface. However, the relative weight for the conditionally-simulated grass surface is always lower than that for the kriged surface.

The estimated parameters, as estimated using WLS, are shown in figure 7.8. The estimates of β_0 and β_1 are similar to those obtained for OLS, but show lower variance. The estimate of σ^2 is approximately 1 in all cases, as expected. The histograms for $\sigma_i^2 = \sigma^2/w_i$ (figure 7.9) show that the values are consistently larger than those obtained for the kriged surface (see table 7.1).

The above analysis emphasises the need to consider the conditionally simulated surface, if an accurate estimate of the variance is required.

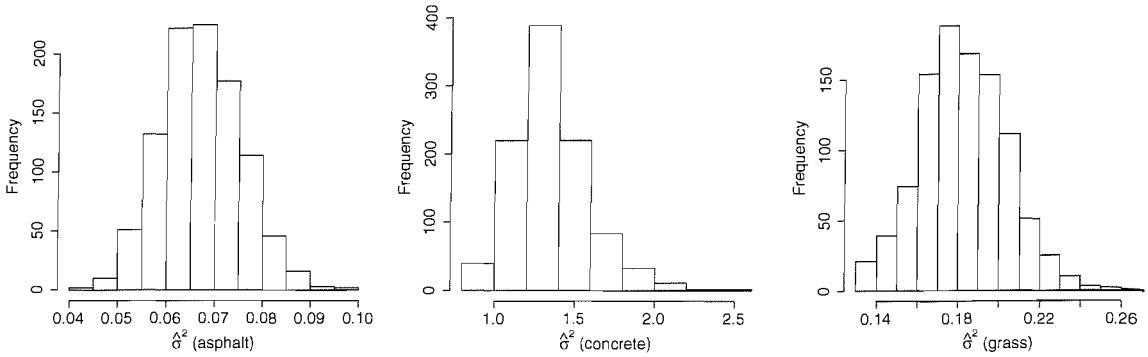


FIGURE 7.9: Histograms showing the estimates of $\sigma_i^2 = \sigma^2/w_i$, obtained for WLS.

Parameter	Estimate	Std. Error	p value
β_0	-2.5454	0.0651	< 0.01
β_1	0.3043	0.0020	< 0.01
σ	1.08		
σ^2	1.02		
R^2	0.98		
	Asphalt	Concrete	Grass
w_i	11.81	0.85	6.94
$w_i/max(w)$	1.0	0.07	0.59
$\sigma_i^2 = \sigma^2/w_i$	0.0861	1.2041	0.1464
σ_i	0.2934	1.0961	0.3826

TABLE 7.4: Table showing the results of the parameter estimation for the ELM regression model after the reflectance measurements have been aggregated using the simple averaging approach. The estimates are for MMR band 1 regressed on ATM band 2.

7.2.2 Simple averaged blocks

In the above analysis (section 7.2.1), the field measurements were aggregated using block kriging and block conditional simulation. However, as discussed in sections 2.2.2.6 and 5.4 there are theoretical and practical problems with using kriged or conditionally simulated surfaces in regression. Hence, the above analysis is repeated for pixel-sized blocks that have been aggregated using the simple averaging approach (section 5.4.1).

The diagnostic regression plots and regression summary statistics are presented in figure 7.10 and table 7.4. It is clear that the weighting has effectively counteracted the heteroskedasticity and the residuals appear normally distributed. The weighting structure differs from the block kriging approach, although that is to be expected.

The confidence and prediction bands for $\mathbf{x_0} = (20, 40, 60)^T$ are shown in tables 7.5 and 7.6 for OLS and WLS respectively. This corresponds to grass, asphalt and concrete respectively. The confidence interval quantifies the uncertainty in the position of the regression line and the prediction interval the uncertainty in making a new prediction. By properly accounting for the heteroskedasticity these can be estimated more accurately. Under WLS

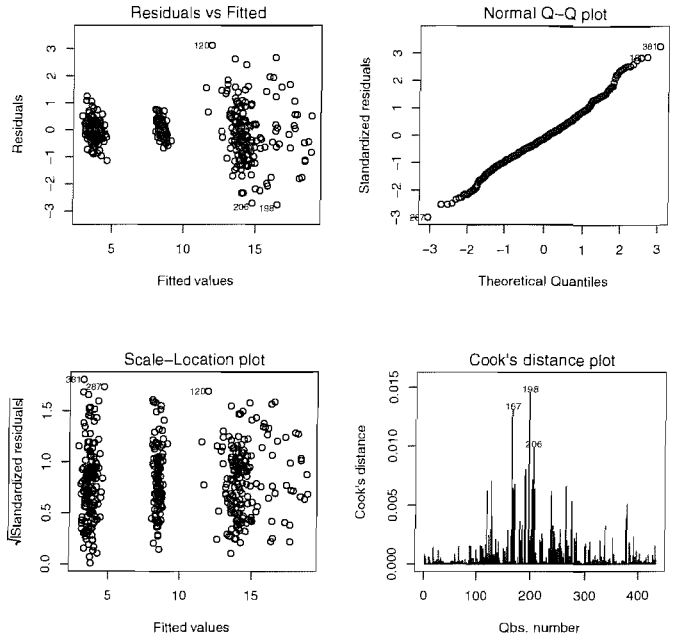


FIGURE 7.10: Diagnostic plots of the residuals for MMR band 1 (simple-averaged blocks) regressed on ATM band 2. The regression is implemented using WLS, where the weights are given as the variance of the OLS residuals for each of three cover types (asphalt, concrete and grass).

x_o	$\hat{y}_0$	Lower CI	Upper CI	Lower PI	Upper PI
20	3.57	3.49	3.64	2.14	5.00
40	9.58	9.53	9.63	8.15	11.01
60	15.59	15.51	15.68	14.16	17.02

TABLE 7.5: Table showing the predicted value of y_0 and the lower and upper 95% confidence (CI) and prediction (PI) intervals. The values of x_0 correspond to grass, asphalt and concrete respectively for MMR band 1 (aggregated using the simple averaging approach) regressed on ATM band 2 using OLS.

the confidence intervals are narrower. Hence there is less uncertainty in the position of the regression line. However, in absolute terms, the confidence intervals are narrow in both cases. As is evident from equation 7.5 the prediction intervals vary from GT to GT, as a function of the weight. For asphalt, the interval is narrower, whereas for concrete it is wider than for the OLS. Hence, if heteroskedasticity is not accounted for the uncertainty in a predicted result may be over- or under-estimated. It should be emphasised that the prediction intervals cannot be calculated for other targets, since the weights are unknown. This is addressed in section 7.3.

This approach was also examined for the other MMR bands (2, 3 and 4) and found to be effective at addressing the effect of heteroskedasticity in the residuals. These results are discussed further in section 7.6.1 and presented in table 7.15.

x_o	$\hat{y}_0$	Lower CI	Upper CI	Lower PI	Upper PI
20	3.55	3.49	3.60	2.79	4.30
40	9.63	9.58	9.69	9.05	10.21
60	15.72	15.59	15.84	13.56	17.88

TABLE 7.6: Table showing the predicted value of y_0 and the lower and upper 95% confidence (CI) and prediction (PI) intervals. The values of x_0 correspond to grass, asphalt and concrete respectively for MMR band 1 (aggregated using the simple averaging approach) regressed on ATM band 2 using WLS.

7.3 Calculating the weights: Strategies for prediction

In section 7.2, tables 7.3 and 7.6 showed the confidence and prediction intervals for the three GTs (asphalt, concrete and grass). However, since the weights were calculated on a per-GT basis, this presented a problem for establishing the weights and associated prediction intervals for a landcover that is not in the estimation data set. Stated simply, it is not known what the weight over such a surface should be. Hence the challenge is to develop an alternative strategy to specify the weights on the basis of a third variable.

7.3.1 Alternative strategies

The magnitude of the residuals, and hence σ_i^2, for a given target may be related to the variance in the image for that target. The first approach was to calculate the variance in the radiance in the 3×3 window surrounding the pixel of interest. This provides the local variance (LV), or local standard deviation (LSD), for each pixel. Alternatively, consideration could be given to using the local maximum difference (LMD) in the 3×3 window surrounding the pixel of interest. The rationale for this approach is that the larger residuals will tend to be associated with pixels that have a larger LV (or LSD or LMD). This relationship is illustrated in figure 7.11, which shows the absolute residuals from kriged surfaces plotted against the LSD for the same pixel. Similar plots are shown in figures 7.12 and 7.13 for the simple-averaged blocks.

Use of the LV opens up three possibilities: calculating the weight on a (i) per-pixel, (ii) per-GT or (iii) per-landcover basis. Calculating the weight on a per pixel basis may lead to unrealistic variation in the weights and subsequent estimate of σ_i^2. This affect will be reduced by calculating the weight on a per target basis. However, in order to extend this method to calculate uncertainty in prediction it is necessary to classify the image by land cover type. This will allow a weight to be tied to a specific land cover type. Hence, the weights that are applied to land cover types will be affected by the classification adopted and decisions made in the classification process.

A further decision needs to be made. Should the weights be established simply by using the LV or average LV (ALV)? Should they be calculated by regressing the OLS residuals against the LV, LSD or LMD? For example, Neter et al. (1996) suggests regressing the

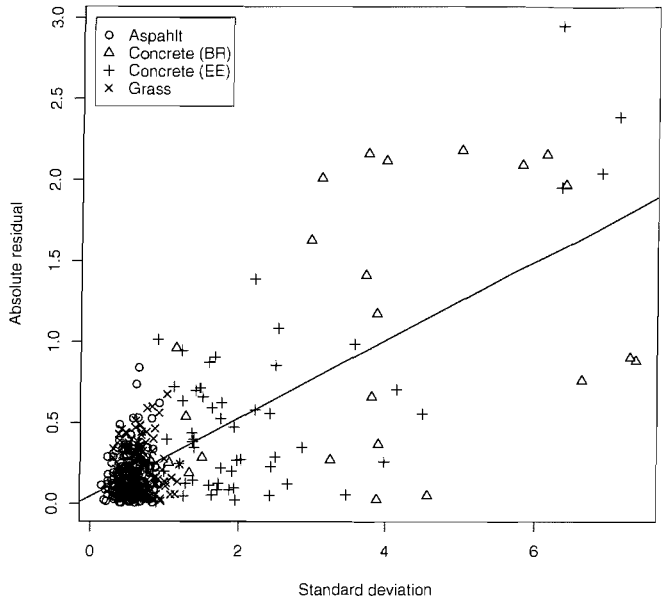


FIGURE 7.11: Plot showing the absolute residual from the OLS regression (MMR band 1 vs ATM band 2) against the standard deviation of the radiance measurements in the surrounding 3×3 window. The MMR measurements have been aggregated using block-kriging.

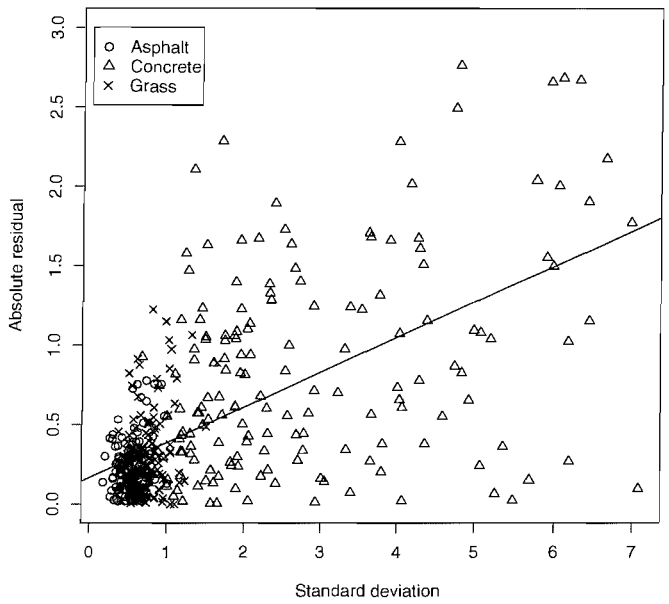


FIGURE 7.12: Plot showing the absolute residual from the OLS regression (MMR band 1 vs ATM band 2) against the standard deviation of the radiance measurements in the surrounding 3×3 window. The MMR measurements were aggregated using the simple averaging approach and all pixels containing *one or more* field measurements are included.

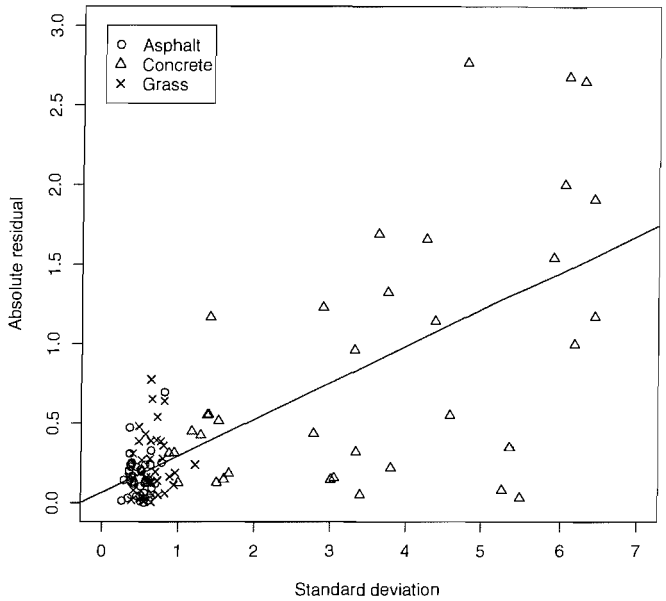


FIGURE 7.13: Plot showing the absolute residual from the OLS regression (MMR band 1 vs ATM band 2) against the standard deviation of the radiance measurements in the surrounding 3 × 3 window. The MMR measurements were aggregated using the simple averaging approach and all pixels containing *two or more* field measurements are included.

absolute OLS residuals or squared OLS residuals against an additional variable? This approach is demonstrated in figures 7.11, 7.12 and 7.13 which show the absolute OLS residuals (for block-kriging and the simple averaging approach) against the LSD for that pixel (see also equations 7.6 and 7.7).

Hence, there are four candidate approaches:

1. Strategy 1: adopt $w_i = 1/LV_i$ on a per-pixel basis.
2. Strategy 2: adopt $w_c = 1/average(LV_i)$ on a per-GT (or per-class) basis.
3. Strategy 3: adopt $w_i = 1/f(LV_i)$ on a per-pixel basis.
4. Strategy 4: adopt $w_c = 1/average(f(LV_i)$ on a per-GT (or per-class) basis.

where $f(LV_i)$ is taken as $|resid_{OLS_i}| = \alpha_0 + \alpha_1 LSD_i$, following equation 7.6.

The analysis presented in this section evaluates the above listed four strategies for establishing a weighting scheme that can be used for prediction. The analysis is performed for MMR band 1 regressed on ATM band 2. Aggregation was performed using the design-based approach. To maintain a large sample size all pixels were retained, even if they only contained one reflectance measurement (see section 5.4.1). The implications of this choice are further discussed later in the chapter.

7.3.2 Image classification

The approaches outlined above discuss calculating the weights on a per-pixel or per-target basis. To implement the latter approach, it was necessary to classify the image into general land cover types which can have a single weight attached to them. This is not straightforward, since it is not clear what such classes should look like. Hence, the decision was made to use broad classes based on maximum-likelihood supervised classification (Campbell 1996, Mather 1999, Richards & Jia 1999, Lillesand et al. 2004), using the ENVI image processing package. The Jeffries-Matusita and Transformed Divergence separability measures (Richards & Jia 1999) were used to test for class separability and seven general classes were established. These were cropped grass, long grass, asphalt, water, trees and scrub, mudflat and concrete. There are areas of low density housing and other buildings in the image. These did not form a separable spectral class, but tended to be subsumed into the other classes. For example, residential areas included components of vegetation, road and concrete. Other areas of the military base tended to be covered by asphalt and concrete (e.g., car parks) and the roofs of large sheds tended to be made from corrugated metal or plastic and were not separable from the concrete class.

It is recognised that, with considerably more effort, a more refined classification might be established which would more accurately represent land cover classes. However, accurate classification of land cover classes was not the objective here. Furthermore, it is not clear whether a more refined classification based on *spectral* separability would yield more appropriate weights.

7.3.3 Strategy 1

The first approach was simply to adopt $w_i = 1/LV_i$ on a per pixel basis. Hence, rather than having a single weight and resulting σ_i^2 for each target there are as many of these as there are pixels in the target. The results are shown in table 7.7. The diagnostic plot for the regression is shown in figure 7.14. Figures showing the histograms for the per-pixel relative weights (figure 7.15) and resulting per-pixel variances (figure 7.16) are provided for each GT.

The diagnostic plot for the regression (figure 7.14) suggests that adopting these weights has substantially reduced the problem of heteroskedasticity. However, a consequence of this approach is that the results are more difficult to interpret since there are many per pixel weights and variances per target. This also leads to wide variability in the per-pixel variance, $\hat{\sigma}_i^2$ (figure 7.15). Hence, it is difficult to draw direct comparisons between these results and the benchmark (table 7.4), although most of the per-pixel relative weights for all three GTs are less than for the benchmark. Finally, figure 7.12 suggests that there is not a direct one-to-one relationship between LV or LSD and the OLS residuals. This latter issue is addressed in strategy 3.

Parameter	Estimate	Std. Error	p value
$\hat{\beta}_0$	-2.5159	0.0578	< 0.001
$\hat{beta}$	0.3031	0.0018	< 0.001
$\hat{\sigma}^2$	0.0.2745		
R^2	0.99		

TABLE 7.7: Table showing the results of the parameter estimation for the ELM regression model after the reflectance measurements have been aggregated using the simple averaging approach. The estimates are for MMR band 1 regressed on ATM band 2. The regression was conducted using WLS, with per-pixel weights given as $w_i = 1/LV_i$ (strategy 1).

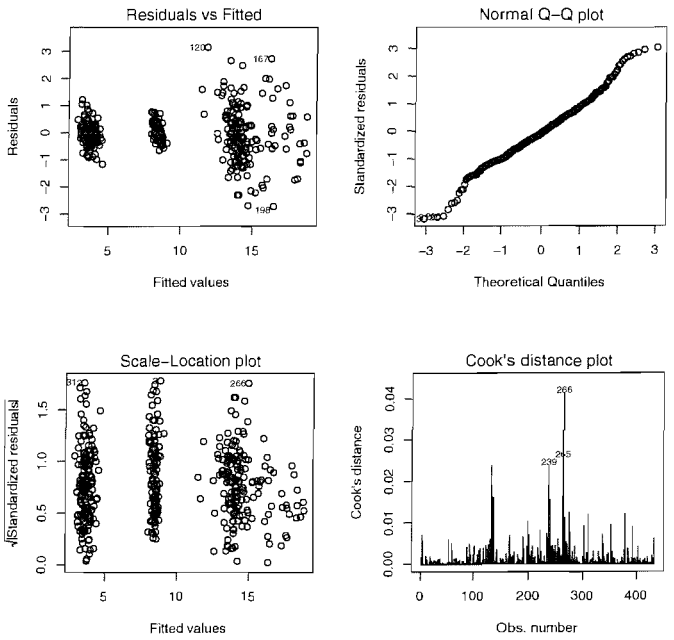


FIGURE 7.14: Diagnostic plots of the residuals for MMR band 1 (simple-averaged blocks) regressed on ATM band 2. The regression is conducted using WLS, with per-pixel weights given as $w_i = 1/LV_i$ (strategy 1).

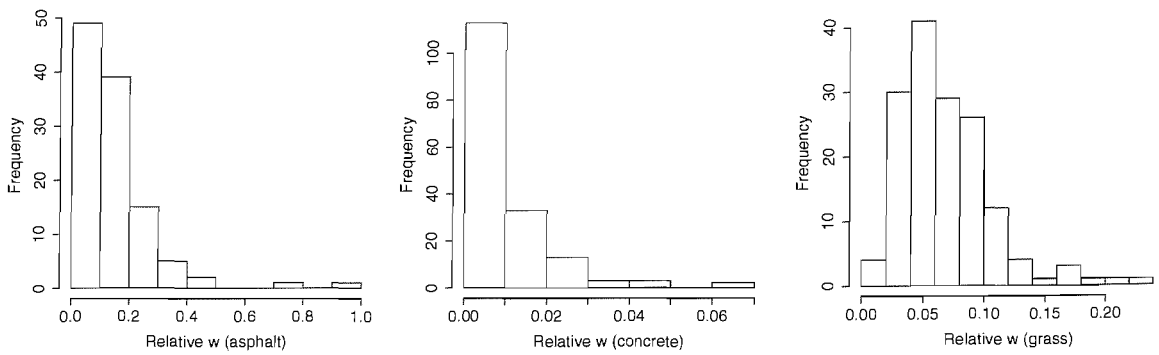


FIGURE 7.15: Histograms showing the per pixel relative weights, given as $w_i/max(w)$ (simple-averaged blocks, strategy 1).

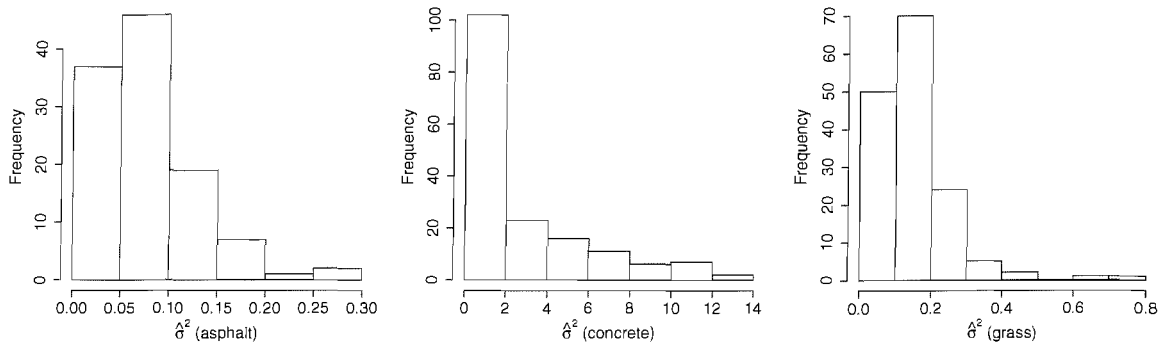


FIGURE 7.16: Histograms showing the per pixel variances, given as $\sigma_i^2 = \sigma^2/w_i$ (simple-averaged blocks, strategy 1).

Parameter	Estimate	Std. Error	p value
$\hat{\beta}_0$	-2.5837	0.0643	< 0.001
$\hat{\beta}_1$	0.3060	0.0020	< 0.001
$\hat{\sigma}^2$	0.2140		
R^2	0.98		
	Asphalt	Concrete	Grass
w_c	3.42	0.0971	1.78
$w_c/max(w)$	1	0.0283	0.5189
$\hat{\sigma}_c^2 = \sigma^2/w_i$	0.0624	2.2036	0.1203
$\hat{\sigma}_c$	0.2499	1.4844	0.3469

TABLE 7.8: Table showing the results of the parameter estimation for the ELM regression model after the reflectance measurements have been aggregated using the simple averaging approach. The estimates are for MMR band 1 regressed on ATM band 2. The regression is conducted using WLS, with *per-GT* weights given as $w_c = 1/average(LV_i)$ (strategy 2).

7.3.4 Strategy 2

The second strategy was to use the target ALV ($w_c = 1/average(LV_i)$) to calculate the average (mean) weight for each target. The results are shown in table 7.8. Table 7.8 also shows the per target weights, relative weights and resultant values for σ_c. The diagnostic plots for the regression are shown in figure 7.17.

The diagnostic plot for the regression (figure 7.17) suggests that adopting these weights has substantially reduced the problem of heteroskedasticity. However, table 7.8 shows that, by comparison to the benchmark for the simple-averaged blocks (table 7.4), the relative weights for concrete and grass are under estimated. A further consequence is that, by comparison to table 7.4, the estimated value of σ_i^2 is slightly under estimated for asphalt and grass and substantially over estimated for concrete.

The above analysis was conducted by averaging over the GTs. For this approach to be of value, it is necessary to be able to predict the class weight for places that were not used as calibration targets. Hence it is necessary to generalise from *targets* to land-

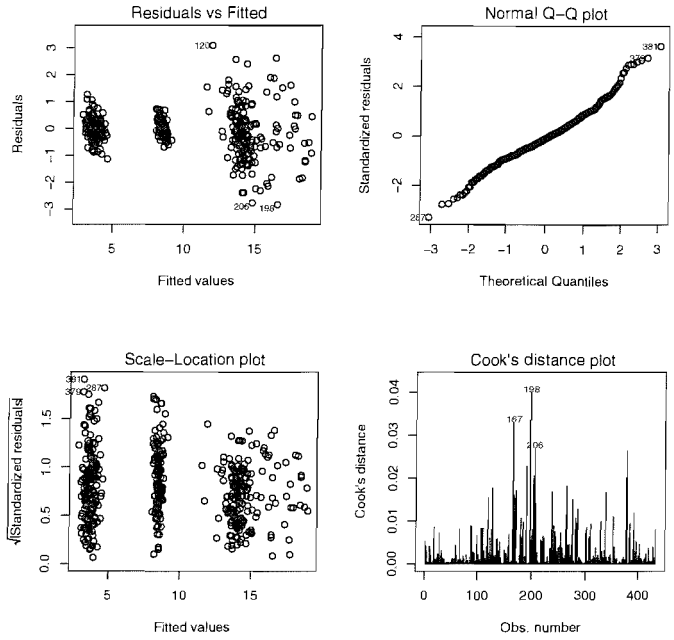


FIGURE 7.17: Diagnostic plots of the residuals for MMR band 1 (simple-averaged blocks) regressed on ATM band 2. The regression is conducted using WLS, with *per-GT* weights given as $w_i = 1/LV_i$ (strategy 2).

cover *classes*. To achieve this the image was classified into eight broad classes, including asphalt, concrete and cropped grass.

The per-class weights were calculated and the regression results are shown in table 7.9 and the diagnostic plots in figure 7.18. This approach is clearly less successful at reducing heteroskedasticity than when the weights are calculated on a per-pixel or per-target basis. It is clear that the relative weights are substantially different to the case where they are calculated for targets, particularly for concrete and grass. By comparison to the reference table 7.4, the estimated value of σ_c^2 is similar for asphalt, slightly underestimated for grass and substantially overestimated for concrete.

These results suggest caution when applying the approach on a per-class basis. In particular, it seems that the within-class variability is large by comparison to the within-target variability. This leads to per-class average LV values that are quite different to the per-target average LV values, with direct consequences for the per-class weights and estimated values for σ_c^2 (see tables 7.8 and 7.9). Furthermore, this suggests that this approach is likely to be sensitive to choices made in image classification.

7.3.5 Strategy 3

The third strategy was to compute the weights as a function of the local variance. This follows the approach adopted by Neter et al. (1996). When following this approach, the OLS residuals were regressed against the LSD ($LSD = \sqrt{LV}$) for the same pixel (see

Parameter	Estimate	Std. Error	p value
$\hat{\beta}_0$	-2.6175	0.0605	< 0.001
$\hat{\beta}_1$	0.3069	0.0021	< 0.001
$\hat{\sigma}^2$	0.0405		
R^2	0.98		
	Asphalt	Concrete	Grass
w_c	0.5438	0.0088	0.4684
$w_c/max(w)$	1.0	0.0161	0.8614
$\hat{\sigma}_c^2 = \sigma^2/w_c$	0.0745	4.6292	0.0865
$\hat{\sigma}_c$	0.2729	2.1515	0.2941

TABLE 7.9: Table showing the results of the parameter estimation for the ELM regression model after the reflectance measurements have been aggregated using the simple averaging approach. The estimates are for MMR band 1 regressed on ATM band 2. The regression is conducted using WLS, with *per-class* weights given as $w_c = 1/average(LV_i)$ (strategy 2).

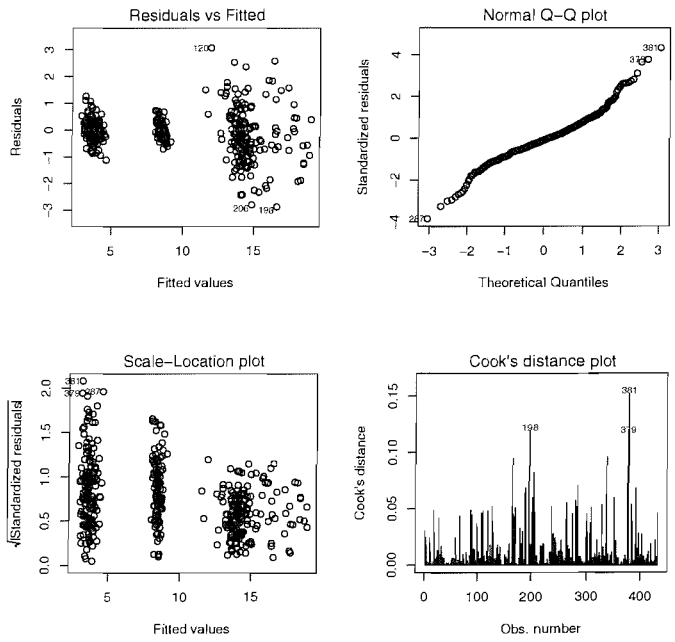


FIGURE 7.18: Diagnostic plots of the residuals for MMR band 1 (simple-averaged blocks) regressed on ATM band 2. The regression is conducted using WLS, with *per-class* weights given as $w_i = 1/LV_i$ (strategy 2).

Parameter	Estimate	Std. Error	p value
$\hat{\beta}_0$	-2.4698	0.0604	< 0.001
$\hat{\beta}_1$	0.3012	0.0018	< 0.001
$\hat{\sigma}^2$	1.5199		
R^2	0.99		

TABLE 7.10: Table showing the results of the parameter estimation for the ELM regression model after the reflectance measurements have been aggregated using the simple averaging approach. The estimates are for MMR band 1 regressed on ATM band 2. The regression is conducted using WLS, with per-pixel weights given as $w_i = 1/f(LV_i)$ (strategy 3).

figure 7.12). The per-pixel weight was then given as $w_i = 1/f(LV_i) = 1/(\alpha_0 + \alpha_1 LSD_i)^2$. The results for the simple-averaged blocks are shown in table 7.10. Diagnostic plots for the regression are shown in figure 7.19. Figures showing histograms for the per-pixel relative weights (figure 7.15) and resulting per-pixel variances (figure 7.16) are provided.

The diagnostic plot for the regression (figure 7.19) suggests that adopting these weights substantially reduced the problem of heteroskedasticity. Regressing the residuals on the LSD ensured that the weights reflect the variability in the residuals, rather than simply the variability in LV. Hence, adopting the regression smoothes some of the variability in the LSD. These combined factors lead to an overall increase in the relative weights compared to strategy 1. The per-pixel variances ($\hat{\sigma}_i^2$) show lower variability than they did for strategy 1. Overall, by comparison to the benchmark shown in table 7.4, the estimated values of σ_i^2 were more accurate than those for strategy 1. The standard errors on the estimated regression coefficients were similar to, but slightly lower than, those for the benchmark.

As with strategy 1, a consequence of strategy 3 is that the results are more difficult to interpret since there are many per-pixel weights and variances per target.

7.3.6 Strategy 4

The final strategy was to average (mean) over the per-pixel weights in strategy 3 on a per-target basis (i.e., $w_c = 1/average(f(LV_c))$). The results are shown in table 7.11. Table 7.11 also shows the per-GT weights, relative weights and resultant values for $\hat{\sigma}_c$. The diagnostic plots for the regression are shown in figure 7.22.

The diagnostic plot for the regression (figure 7.22) suggests that adopting these weights has substantially reduced the problem of heteroskedasticity. By comparison to strategy 2, table 7.8 shows that the relative weights accurately reflect those for the benchmark (table 7.4). This accuracy is also reflected in the estimated values of σ_c^2 for each target.

Previous analysis for strategy 4 was conducted by averaging over the *GTs*. However, for this approach to be of value, it is necessary to be able to predict the class weight for places

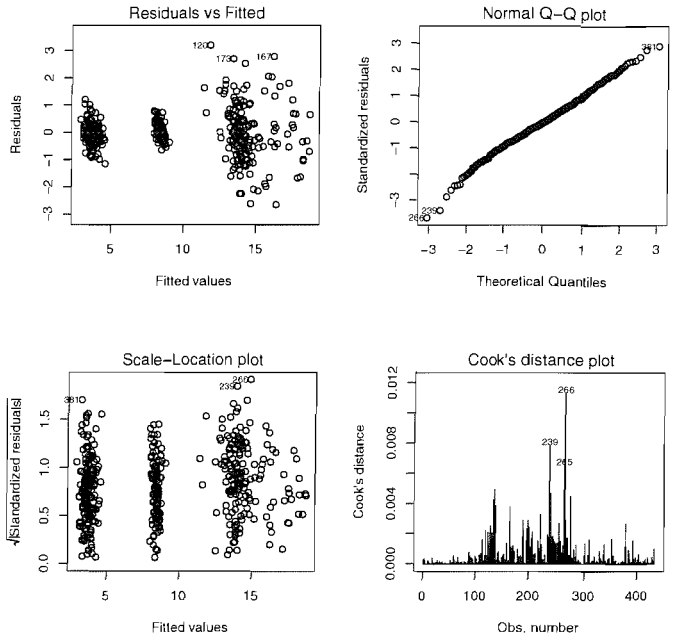


FIGURE 7.19: Diagnostic plots of the residuals for MMR band 1 (simple-averaged blocks) regressed on ATM band 2. The regression is conducted using WLS, with per-pixel weights given as $w_i = 1/f(LV_i)$ (strategy 3).

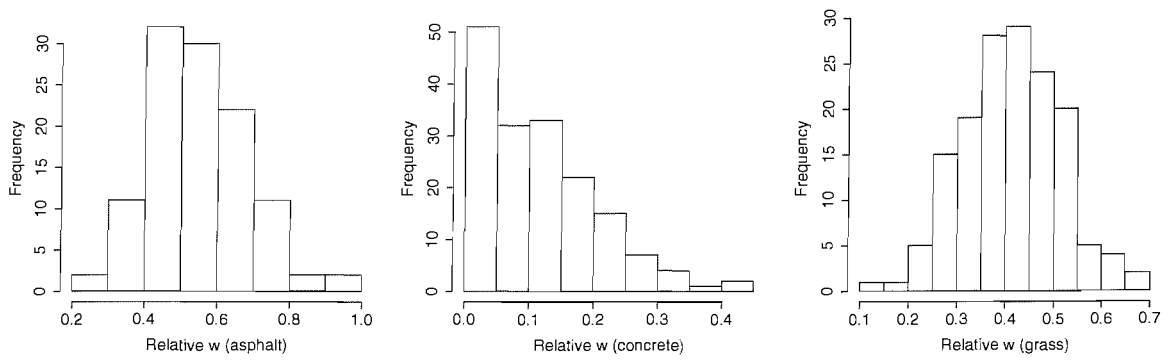


FIGURE 7.20: Histograms showing the per pixel relative weights, given as $w_i/max(w)$ (strategy 3).

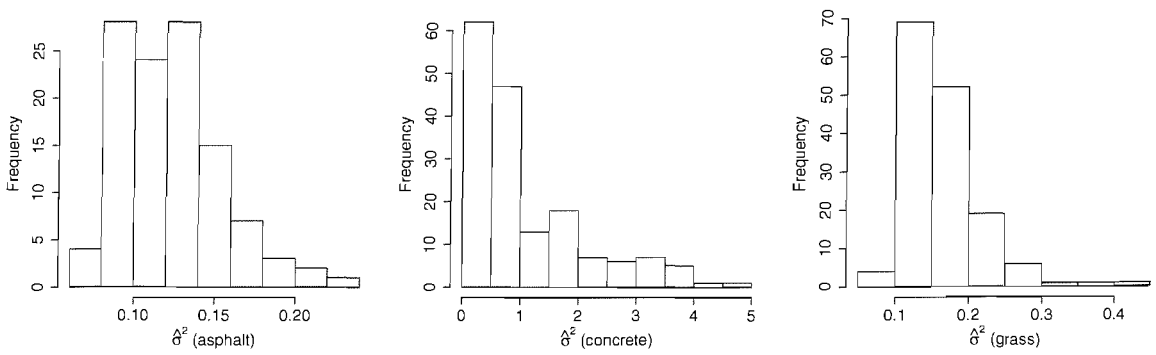


FIGURE 7.21: Histograms showing the per pixel variances, given as $\sigma_i^2 = \sigma^2/w_i$ (strategy 3).

Parameter	Estimate	Std. Error	p value
$\hat{\beta}_0$	-2.5159	0.0642	< 0.001
$\hat{\beta}_1$	0.3034	0.0020	< 0.001
$\hat{\sigma}^2$	1.3554		
$\hat{\sigma}$	1.1642		
R^2	0.98		
	Asphalt	Concrete	Grass
w_c	12.3862	1.3583	9.1603
$w_c/max(w)$	1.0	0.1091	0.7396
$\hat{\sigma}_c^2 = \hat{\sigma}^2/w_c$	0.1094	1.0026	0.1480
$\hat{\sigma}_c$	0.3308	1.0013	0.3847

TABLE 7.11: Table showing the results of the parameter estimation for the ELM regression model after the reflectance measurements have been aggregated using the simple averaging approach. The estimates are for MMR band 1 regressed on ATM band 2. The regression is conducted using WLS, with *per-GT* weights given as $w_c = 1/average(f(LV_i))$ (strategy 4).

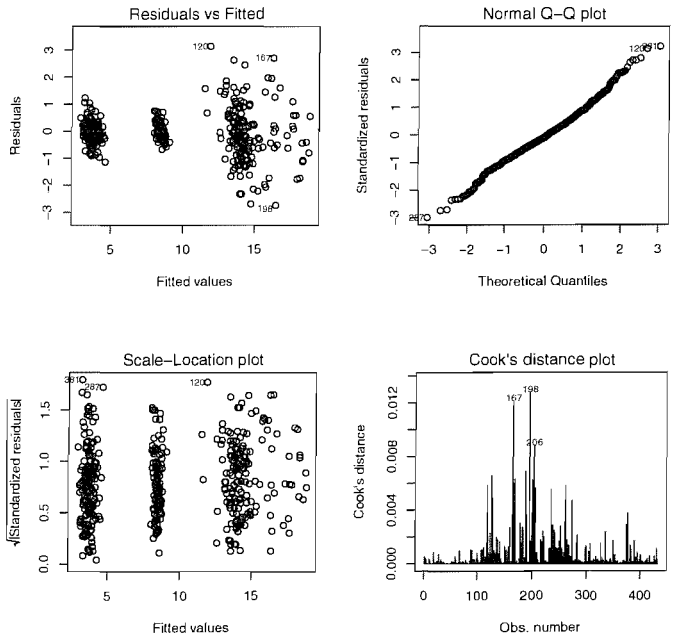


FIGURE 7.22: Diagnostic plots of the residuals for MMR band 1 (simple-averaged blocks) regressed on ATM band 2. The regression is conducted using WLS, with *per-GT* weights given as $w_c = 1/average(f(LV_i))$ (strategy 4).

Parameter	Estimate	Std. Error	p value
$\hat{\beta}_0$	-2.5918	0.0603	< 0.001
$\hat{\beta}_1$	0.3060	0.0021	< 0.001
$\hat{\sigma}^2$	0.4311		
$\hat{\sigma}$	0.6566		
R^2	0.98		
	Asphalt	Concrete	Grass
w_c	5.3393	0.1615	4.7659
$w_c/max(w)$	1.0	0.0302	0.8926
$\hat{\sigma}_c^2 = \sigma^2/w_i$	0.0807	2.6754	0.0905
$\hat{\sigma}_c$	0.2842	1.6357	0.3008

TABLE 7.12: Table showing the results of the parameter estimation for the ELM regression model after the reflectance measurements have been aggregated using the simple averaging approach. The estimates are for MMR band 1 regressed on ATM band 2. The regression is conducted using WLS, with *per-class* weights given as $w_c = 1/average(f(LV_i))$.

that were not used as calibration targets. Hence, as with strategy 2 it was necessary to generalise from *GTs* to land-cover *classes*.

The per-class weights were calculated and applied for simple-averaged blocks. The diagnostic plots are shown in figure 7.23 and the regression results in table 7.12. The approach has not been effective at removing the effects of heteroskedasticity. Furthermore, this approach only gives a small increase in accuracy compared to the case when strategy 2 is applied on a per-class basis (table 7.9) and is inaccurate compared to the benchmark (table 7.4). As with strategy 2, these results suggest caution when applying the approach on a per-class rather than per-target basis.

7.3.7 Summary

The above analysis showed that using the LV reflects the general ordering of the weights – i.e., asphalt should be more highly weighted than grass, which should be more highly weighted than concrete. Strategies 1 and 2 substitute the LV for the weight directly, whereas Strategies 3 and 4 predict the weight as a function of the LV. Hence, the first two strategies assume a direct relationship between the weight and the LV, whereas the latter two model this relationship.

For strategies 1 and 2 the order of the weights was correct, but the actual values of the relative weights (and consequent estimates of σ_i^2 or σ_c^2) were inaccurate relative to the benchmark data set. However, the results were more accurate for strategies 3 and 4. This additional accuracy may be attributed to the explicit modelling of the weights, rather than simply substituting the LV.

The analysis raised some debate as to whether the weights and consequent values of σ_i^2 should be calculated on a per-pixel, per-GT or per-class basis. The per-GT or per-class

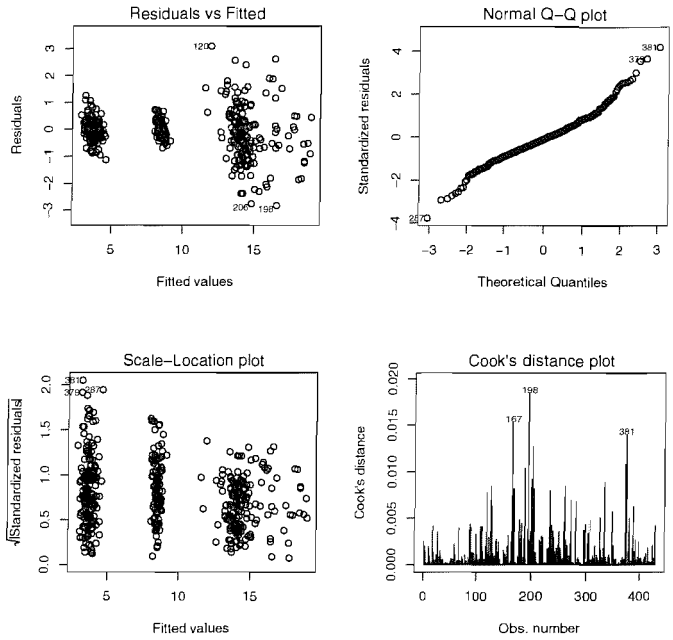


FIGURE 7.23: Diagnostic plots of the residuals for MMR band 1 (simple-averaged blocks) regressed on ATM band 2. The regression is conducted using WLS, with *per-class* weights given as $w_c = 1/average(f(LV_i))$ (strategy 4).

approach should mitigate against large random variations in the weights of individual pixels. However, this needs to be balanced against the fact that it smoothes out within-GT or within-class variability. Indeed, it might be argued that averaging over a GT or class effectively increases the support of the weight. This might be considered to be undesirable, given that prediction on a per-pixel basis is required.

A serious concern when calculating weights on a per-class basis arose when transferring from *per-GT* to *per-(land cover) class* weights. The resulting per-class relative weights (and consequent estimates of per-class σ_c^2) were substantially different from the per-GT relative weights and σ_c^2 estimates. These implied that the within-class variability led to the prediction of weights that were substantially different from the associated per-GT weights. Furthermore, the per-class approach is likely to be sensitive to the classification scheme adopted. On the basis of this evidence, it is not possible to provide generalised weights for landcover classes. Hence, it is recommended that the weights should be predicted on a per-pixel basis using strategy 3.

The weighting schemes discussed above were also investigated for MMR bands 2, 3 and 4. Broadly, strategies 2 and 4 presented the same problems when calculating the weights on a per-class basis. For strategies 1 and 3 and the other visible wavebands (MMR bands 2 and 3), similar results were obtained to band 1. Hence, strategy 1 did reduce the impact of heteroskedasticity, whereas strategy 3 practically eliminated the effect.

Despite the success for the visible wavebands, strategies 1 and 3 were less effective at reducing the effect of heteroskedasticity for the NIR band (MMR band 4). Figure 7.24

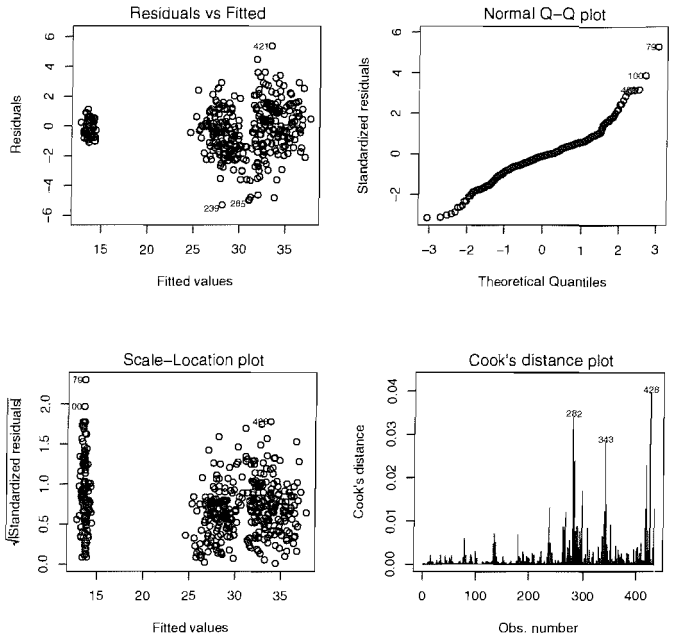


FIGURE 7.24: Diagnostic plots of the residuals for MMR band 4 (simple-averaged blocks) regressed on ATM band 7. The regression is conducted using WLS, with per-pixel weights given as $w_i = 1/LV_i$ (strategy 1).

shows the diagnostic plots for strategy 1, whereas figure 7.25 shows the diagnostic plots for strategy 3. The standardised residuals in figure 7.25 show strategy 1 was effective at reducing heteroskedasticity, although the QQ plot shows that these residuals depart from normality. Strategy 3 (figure 7.25) was more effective in this latter respect, but not effective at reducing the effects of heteroskedasticity. Alternative regression models were examined as alternative implementations of strategy 3, although none was substantially more effective.

In conclusion, strategy 3 emerged as the preferred approach for the visible wavebands. However, there was no effective strategy for the NIR waveband.

7.4 Dealing with correlation amongst the residuals

In the introductory section of this chapter it was emphasised that ordinary least squares (OLS) regression modelling assumes that the model residuals are independent and identically distributed (iid). Sections 7.1, 7.2 and 7.3 addressed the issue of heteroskedasticity. This section addresses spatial correlation in the residuals. This was addressed in chapter 2 and particularly in sections 2.1.4, 2.2.2.4, 2.2.2.5 and 2.4. The methodology presented in this section builds on the maximum likelihood approach to regression modelling. Analysis is performed first using simulated data, before using the methodology on real data gathered for implementing the ELM. Simulated data are used initially since they allow evaluation of the accuracy of the methodology when the results are known.

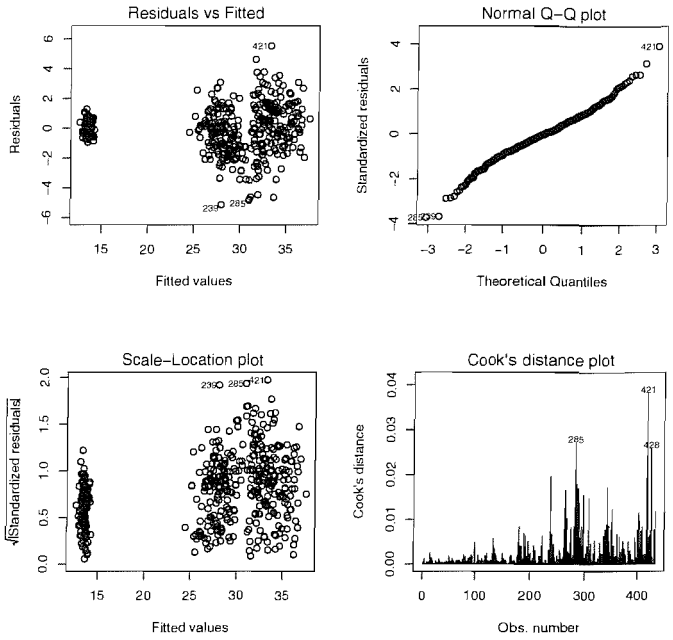


FIGURE 7.25: Diagnostic plots of the residuals for MMR band 4 (simple-averaged blocks) regressed on ATM band 7. The regression is conducted using WLS, with per-pixel weights given as $w_i = 1/f(LV_i)$ (strategy 3).

As noted previously, where the covariance matrix of the residuals, $\mathbf{\Sigma}$ is known $\beta = (\beta_0, \beta_1)^T$ is estimated using generalised least squares (GLS) as:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{\Sigma}^{-1}\mathbf{y} \tag{7.8}$$

If the residuals are homoskedastic, $\mathbf{\Sigma}$ can be replaced by $\mathbf{R}$, a correlation matrix such that $\mathbf{\Sigma} = \sigma^2\mathbf{R}$. Alternatively, if there is heteroskedasticity but no correlation, $\mathbf{R}$ can be replaced by the weights matrix $\mathbf{W}$. The covariance matrix of $\hat{\beta}$ is then given as

$$Cov(\hat{\beta}) = (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} \tag{7.9}$$

The presence of correlation amongst the residuals can have various related effects on the regression model. First, as shown in equation 7.8 the correlation structure can directly influence the estimate of β. Second, the correlation structure can affect the estimate of σ^2, estimated with ML as (Harris & Johnson 1996, Lark 2000):

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\beta)^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\beta) \tag{7.10}$$

Since the correlation structure affects the estimate of $Cov(\hat{\beta})$ (equation 7.9), even if the GLS estimator of β is similar to the OLS estimator, $Cov(\hat{\beta})$ will differ. That has implications for quantifying uncertainty in estimation and prediction, for specifying confidence and prediction intervals and for hypothesis testing. In particular, the OLS estimators tend to underestimate σ^2 and yield inaccurate confidence intervals (Harris &

Johnson 1996, Ver Hoef et al. 2001, Lark & Cullis 2004). This can affect the outcome of hypothesis testing as well as choices as to which covariates should be included in the model. If correlation is not accounted for then it can be conflated with a trend, leading to the inclusion of spurious covariates in a regression model (Ver Hoef et al. 2001, Hengl et al. 2004). Finally, not accounting for correlation can lead to inaccurate estimation of the confidence and prediction intervals.

The above discussion assumed that $\mathbf{\Sigma}$ is known. In practice this is rarely the case and it is necessary to estimate the elements of $\mathbf{\Sigma}$. This can be achieved through embedding a geostatistical variogram model to account for the spatial correlation which is then used to predict the elements of $\mathbf{\Sigma}$. A framework is adopted that iterates between calculation of $\hat{\beta}$ and $\mathbf{\Sigma}$ to maximise the log-likelihood function for the Gaussian model:

$$-\frac{1}{2}(\log|\mathbf{\Sigma}| + (\mathbf{y} - \mathbf{X}\beta)^T\mathbf{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\beta)) \tag{7.11}$$

Within the Gaussian ML framework, instead of assuming that the regression residuals are $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ the residuals are modelled as being multivariate-Gaussian, $\varepsilon \sim MVN(0, \mathbf{\Sigma})$. Mathematically, this approach has the same setup as the ML implementation of model-based geostatistics, discussed in section 2.2.2.5. However, as stated in section 2.4 the emphasis is on *regression* rather than *interpolation.*

Previous developments of this ML approach (Cook & Pocock 1983, Mardia & Marshall 1984, Lark 2000) proceeded by setting $\mathbf{\Sigma} = \sigma^2\mathbf{R}$. $\mathbf{R}$ is a correlation matrix, with the off diagonal elements by $r_{ij} = f(\phi, \mathbf{h}_{ij})$, where $\mathbf{h}_{ij}$ is the lag between i and j and $f(\cdot,\cdot)$ is an authorised correlation (or variogram) function. Typically the exponential or spherical covariance function is adopted, requiring the estimation of the range parameter, ϕ. The regression parameters, for a given value of ϕ, are given as:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{R}^{-1}\mathbf{y} \tag{7.12}$$

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\beta)^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\beta) \tag{7.13}$$

$$Cov(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1} \tag{7.14}$$

and the log-likelihood reduces to:

$$-\log|\mathbf{R}| - n\log\sigma^2 \tag{7.15}$$

This approach was initially proposed in the mid-1980s (Cook & Pocock 1983, Mardia & Marshall 1984), and has been widely cited in the statistical and epidemiological literature. However, there has only been limited uptake in remote sensing and other environmental fields. Lark (2000) demonstrated the above approach for simulated imagery and found that it gave more accurate estimates of σ^2 than OLS, with OLS tending to underestimate σ^2. This had important consequences for uncertainty analysis and for tests on the regression

parameters. However, Lark did not examine the accuracy of the estimates of ϕ, which is clearly important since it affects the elements of $\mathbf{\Sigma}$. Furthermore, this model does not specify a nugget component.

The above approach was revised by Lark & Cullis (2004), who specified:

$$\begin{aligned} \mathbf{\Sigma}_{ij} &= \sigma^2 \nu f(\phi, \mathbf{h}_{ij}), i \neq j \\ &= \sigma^2, i = j \end{aligned} \tag{7.16}$$

The parameter ν is defined as

$$\nu = \frac{\tau_s}{\tau_n + \tau_s}$$

which allows incorporation of a nugget effect, since τ_s is the partial sill and τ_n is the nugget component. Hence $\sigma^2 = \tau_s + \tau_n$. The estimates and covariance function of β are then given as in equations 7.8 and 7.9. The variogram parameters can then be estimated by interactively maximising the likelihood.

As discussed in section 2.2.2.5, ML is known to yield biased estimates of σ^2. This arises since, although equation 7.13 takes account of the correlation structure, the denominator is n. This is known to be a biased estimator since it does not take into account the loss of degrees of freedom caused by estimating β (Neter et al. 1996). Hence, Harris & Johnson (1996) recommend adjusting for the degrees of freedom, as follows:

$$\tilde{\sigma}^2 = \frac{n\hat{\sigma}^2}{n-p} \tag{7.17}$$

where n is the number of data points used in the regression model and p is the number of model coefficients. However, a more widely cited approach is that of residual (or restricted) maximum likelihood (REML) (Cressie 1993, Diggle et al. 2003) as advocated by Lark & Cullis (2004). REML proceeds by maximising the conditional likelihood:

$$l(\sigma^2, \nu, \phi | \hat{\beta}, \beta) \propto -\frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\log|\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{Y}^T\mathbf{\Sigma}^{-1}(\mathbf{I}-\mathbf{Q})\mathbf{y} \tag{7.18}$$

where

$$Q \equiv \mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}$$

As noted previously, although REML is widely used in geostatistics (Cressie 1993), there is not a consensus that it should be preferred to ML (Diggle et al. 2003). Furthermore, the fact that the residuals are correlated effectively leads to a further decrease in the number of degrees of freedom (Ver Hoef et al. 2001). Indeed, this loss of information presents a problem for estimation (it reduces the information content) but an advantage for prediction (since the predicted location will be correlated with the surrounding data points) (Ver Hoef et al. 2001).

Lark & Cullis (2004) demonstrated the use of the REML approach for data gathered in the Swiss Jura, but do not provide an empirical evaluation against simulated data in

a remote sensing context. Furthermore despite their recommendations, neither Lark & Cullis (2004) or Diggle et al. (2003) provide an empirical evaluation of ML *vs.* REML.

This section evaluates the ML and REML approaches using simulated data. This extended the approach of Lark (2000) since it examined the accuracy of the estimation of ϕ and ν as well as conducting a comparison of the ML and REML approaches.

7.4.1 Experiments with simulated imagery

The utility of the ML and REML approaches was investigated using simulated imagery, following the approach of Lark (2000). The images were created on a uniform square grid of 249 rows by 249 columns. First, a random standard normal variable $\mathbf{X}$ was created on the square grid, as illustrated in figure 7.26. Next, an image of errors (ε) were created, with variance, $\sigma^2 = 1$, partitioned into a nugget term, $\tau_n = 0.5$ and a partial sill, $\tau_s = 0.5$, i.e., $\nu = 0.5$. Finally, the dependent variable, y, was generated where, $\beta = (1, 1.2)^T$, hence:

$$y_i = 1 + 1.2x_i + \varepsilon_i \tag{7.19}$$

This is the same set-up as that used by Lark (2000), although he set $\nu = 0.95$, but omitted the nugget effect from his model.

The errors were simulated using a spherical variogram model. When estimating the model parameters (whether by ML or REML), it is necessary to choose an authorised model for $f(\phi, \mathbf{h}_{ij})$ (equation 7.16). The smooth, differentiable nature of the exponential model makes it well suited to ML techniques (Diggle et al. 2003, Lark & Cullis 2004). However, the spherical model is so widely used in geostatistics that it is reasonable to examine it in this context. A choice between the spherical and exponential model can then be made on the basis of the maximized likelihood and the Akaike Information Criterion (AIC) . It may be noted that, in any given application, the spherical or exponential model may not reflect the underlying data. In such cases, other variogram models could be investigated. However, it is also useful to consider the implication of choosing the *wrong* variogram model and whether this has implications for parameter estimation.

Six images of $\mathbf{X}$, ε and the corresponding images of $\mathbf{y}$ were created for $\phi = 1, 2, \ldots, 6$. The image of ε was created using unconditional simulation (Sequential Gaussian Simulation). The images are shown in figures 7.27 and 7.28. For each image two sub-samples of 100 data were obtained. The first sub-sample was gathered using simple random sampling (SRS) and second using spatially structured sampling (SS) on a 10×10 grid with 1 pixel spacing, as shown in figure 7.29. This was repeated 100 times, without replacement, as shown in figure 7.30. These repeat samples allowed investigation of the accuracy of the parameter estimates.

Having simulated the images and drawn samples from them, as described above, the model parameters were estimated using OLS, ML and REML. The results are discussed

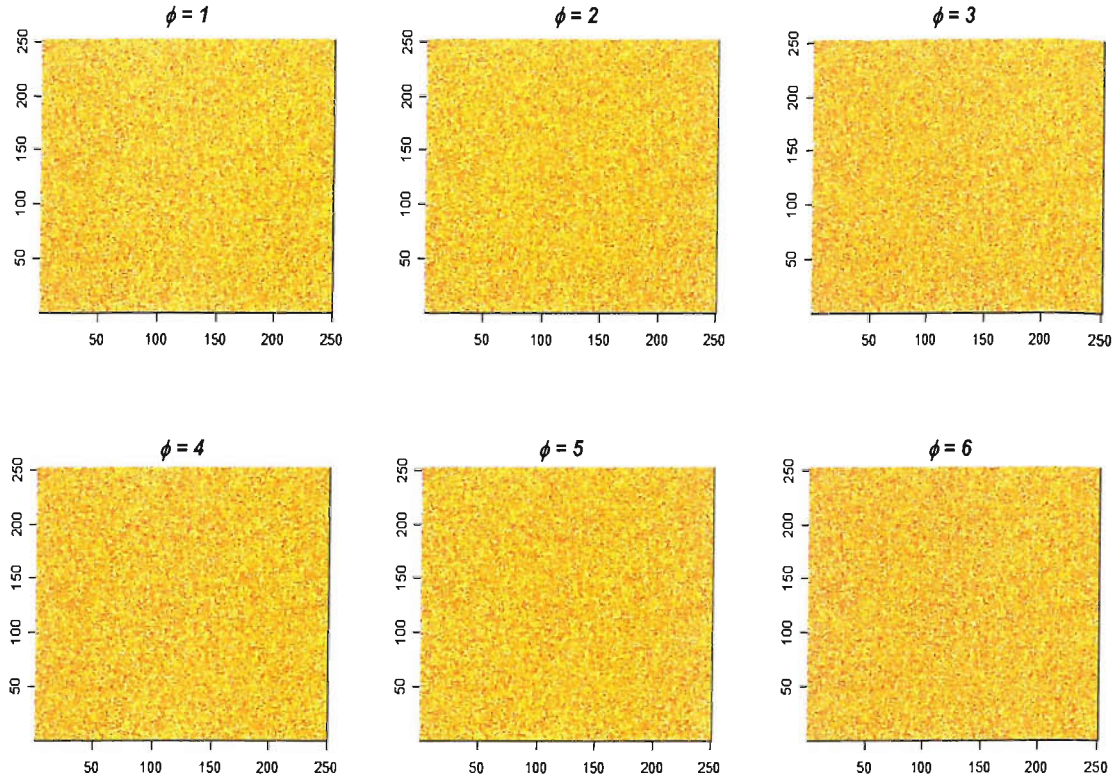


FIGURE 7.26: Images of x used in equation 7.19 and simulated as $X \overset{iid}{\sim} N(0,1)$. One image is duplicated for $\phi = 1, 2, \ldots, 6$.

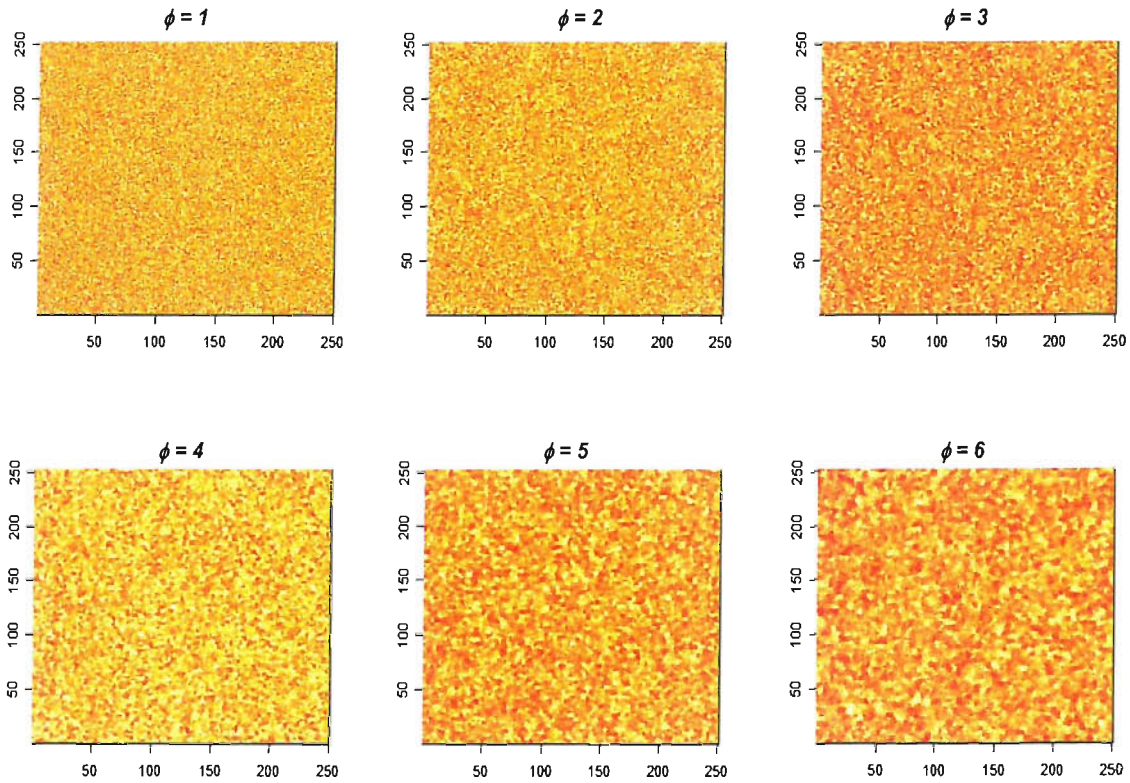


FIGURE 7.27: Unconditionally simulated images (one realisation) for the error term, ε. They were modelled using a spherical model with $\sigma^2 = 1$, $\nu = 0.5$ and $\phi = 1, 2, \ldots, 6$.

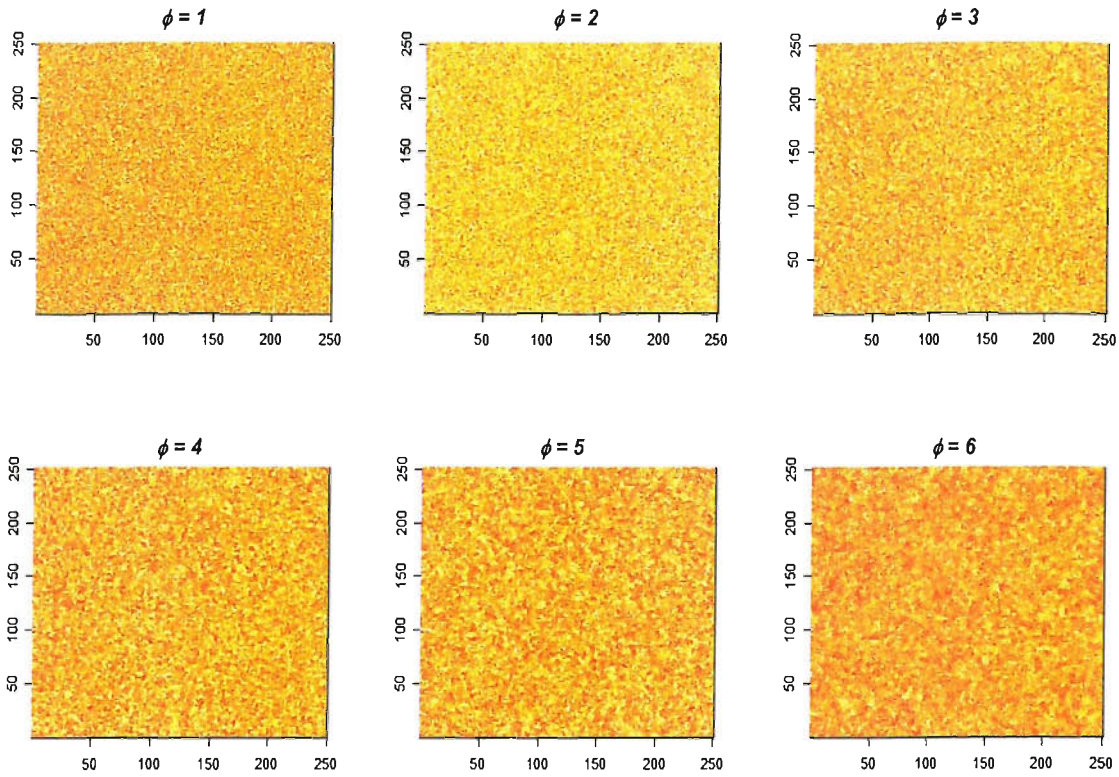


FIGURE 7.28: One realisation of $y_i = 1 + 1.2x_i + \varepsilon_i$, where $X_i \sim N(0, 1)$ and the ε are as presented in figure 7.27.

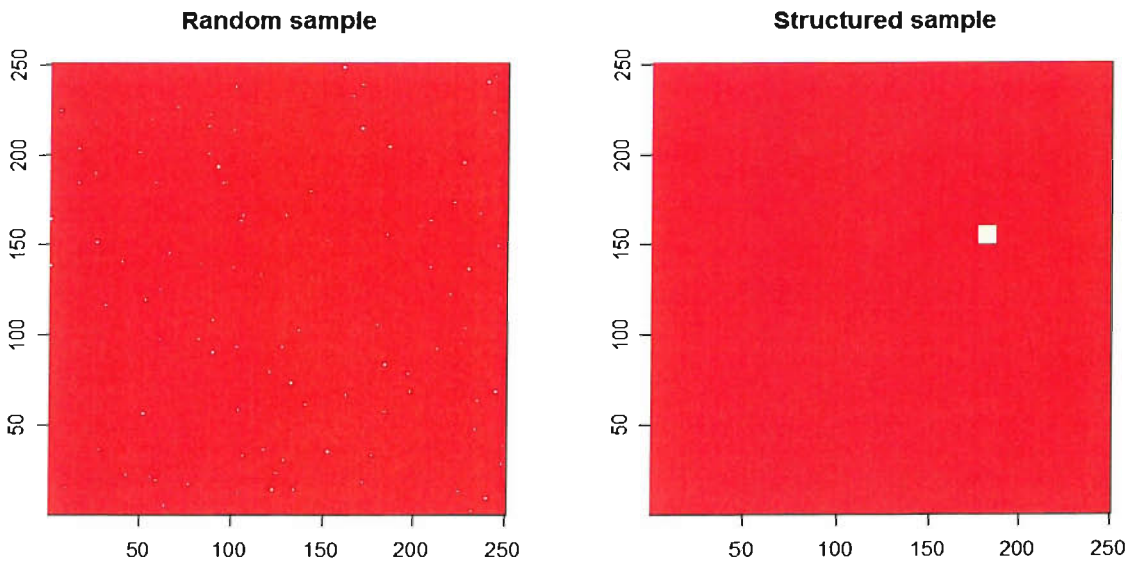


FIGURE 7.29: Image showing one realisation of a random and structured sample of 100 measurements.

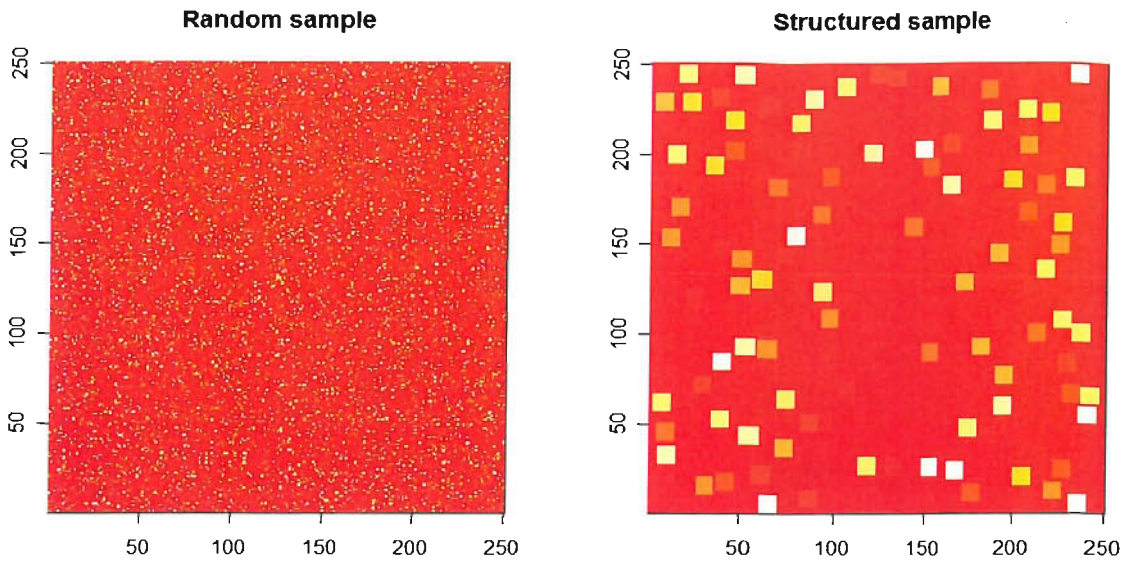


FIGURE 7.30: Image showing 100 realisations of a random and structured sample of 100 measurements. The colours indicate different realisations

below.

7.4.1.1 Results

Figure 7.31 shows the estimated value of σ plotted against the range, ϕ, of the error variable, ε. Recall that six different images of ε were simulated for $\phi = 1, 2, \ldots, 6$. These show the average estimate of σ (for the 100 samples), together with 95% confidence intervals on the mean as well as the first and third quartiles. Figure 7.31 (a, c, e) shows the estimated value of σ obtained using random sampling. It is clear that, under simple random sampling ML or REML offer no advantage over OLS for providing unbiased estimates of σ.

Under systematic sampling figure 7.31 (b) shows the estimates of σ are negatively biased when estimated using OLS. This is particularly clear when the range of the underlying error variable was large ($\phi = 4, 5, 6$), compared to the size of area sampled (10×10). When ϕ is small relative to the size of the area sampled, many of the residuals will be uncorrelated. However, when ϕ is large a greater proportion of the residuals will be correlated. Hence, when ϕ is relatively large, the effect of the spatial correlation is greater (Lark 2000).

The ML and REML models were then implemented by adopting a spherical model for equation 7.16. The resulting estimates of σ are shown in figure 7.31 (d and f). It is interesting to note that ML (d) does not seem to offer a substantial improvement over OLS. However, REML (f) does substantially reduce the bias. This differs from the result found by (Lark 2000), who found that adopting the ML approach eliminated bias in σ although he set $\nu = 0.95$, rather than 0.5. Lark's results were replicated when the algorithms were implemented for this analysis. Hence, the magnitude of ν may affect the

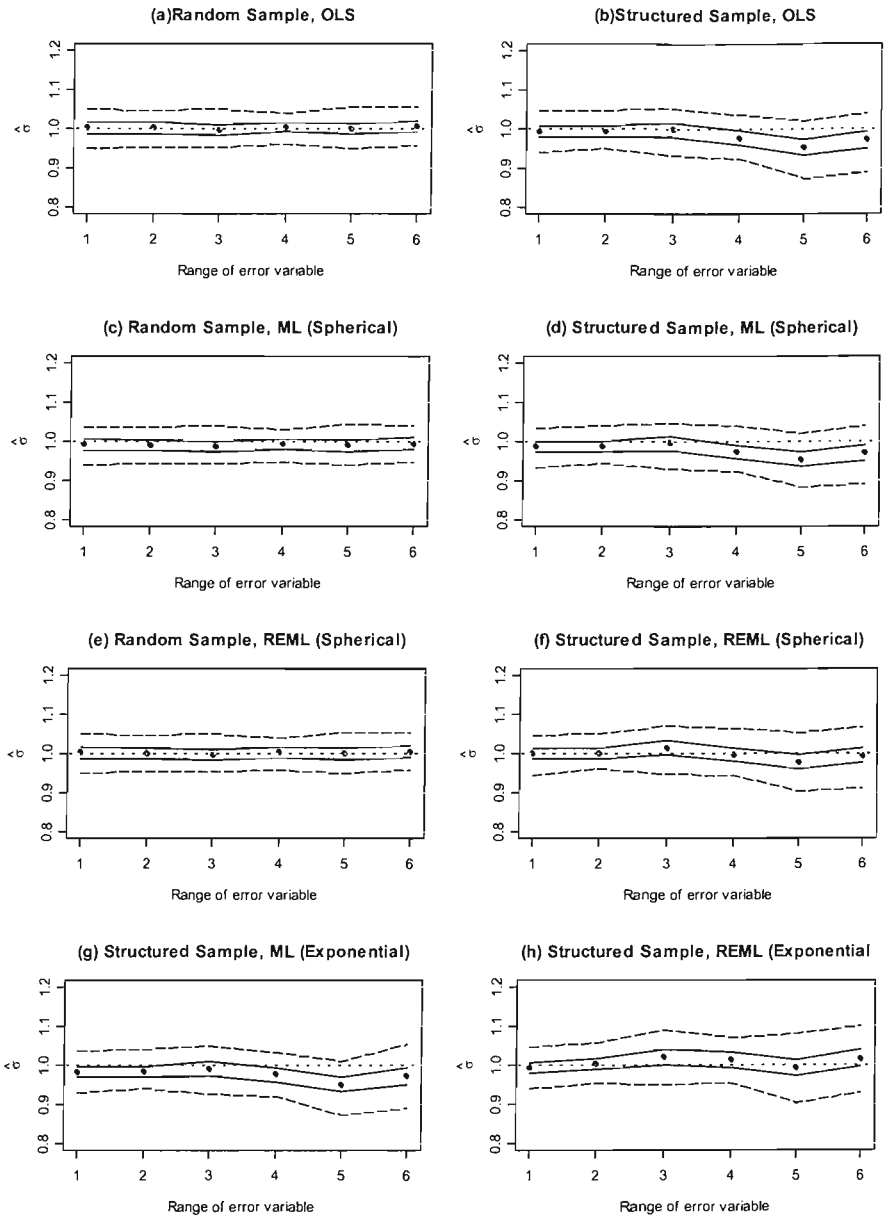


FIGURE 7.31: Estimated value of σ of a given range of the error variable. The points show the mean estimate of σ. Solid lines show the 95% confidence intervals on the mean and the dashed line show the upper and lower quartiles.

accuracy of the ML approach. Finally, when the spatial covariance is modelled using an exponential model figure 7.31 (g and h) a similar pattern is observed to that found for the spherical model. This shows that the estimation of σ^2 was robust to choice of variogram model.

Boxplots showing the estimates of ϕ are shown in figures 7.32 and 7.33 for the random and structured samples respectively. Under random sampling (figure 7.32), the estimates were highly variable and inaccurate.The results for the structured sample (figure 7.33) are now considered in more detail. First the spherical model was considered (figure 7.33 (a and b)). Both ML and REML gave unbiased estimates; however, the REML estimates tended

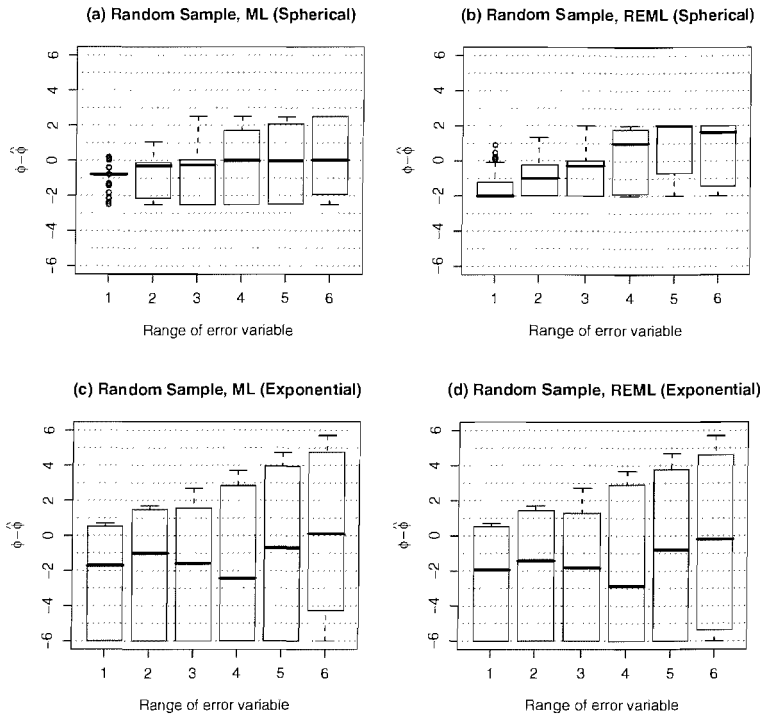


FIGURE 7.32: Boxplots showing $\phi - \hat{\phi}$ for a given range of the error variable, ε. For the exponential model the *effective* range is shown. The results are shown for simple random sampling.

to be less *precise* than the ML estimates, in that they show greater variability. Indeed, practical experience found numerical instability in the REML estimates, contributing to variability in those estimates. The extent of that variability is masked in these boxplots. Greater variability in the REML estimates was also found for $\hat{\sigma}$ (figure 7.31).

The results for the exponential model are shown in figure 7.33 (c and d). The range given is the effective range ($\phi/3$) and should not be expected to correspond exactly with the range of the error variable for the underlying surface (which was modelled using a spherical variogram). For both ML and REML the estimates of the range approximate the range of the underlying variogram. However, the precision of the estimated values was lower than where the spherical model was adopted. As before, greater instability was found when implementing the REML procedure.

Boxplots showing the estimates of ν are shown in figures 7.34 and 7.35 for the random and structured samples respectively. As with ϕ, the estimates under random sampling (figure 7.34) were highly variable and inaccurate. For the structured sample and the spherical model (figure 7.35 (a)) the estimates were highly variable and were inaccurate for $\phi = 1$ and 2. However, for large values of ϕ the accuracy of estimation increased. A similar pattern was observed for the REML estimates (figure 7.35 (b)) and the spherical model. However, as with the estimates of ϕ, the boxplots do not show the instability and variability of the REML estimates.

The results for the exponential model are shown in figure 7.35 (a and b). As with the

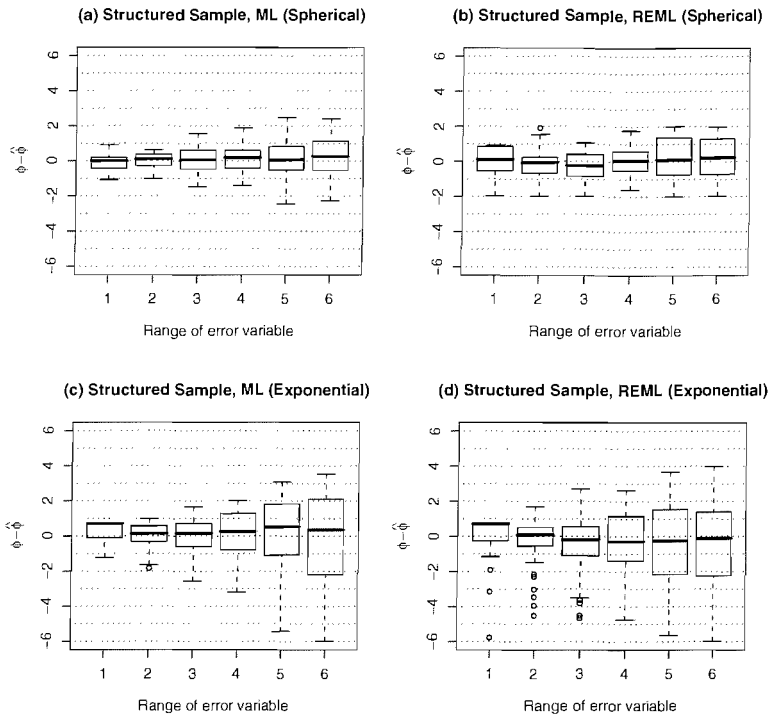


FIGURE 7.33: Boxplots showing $\phi - \hat{\phi}$ for a given range of the error variable, ε. For the exponential model the *effective* range is shown. The results are shown for structured sampling on a 10×10 grid.

spherical model, the estimates were highly variable and inaccurate for $\phi = 1$ and 2. At larger values of ϕ the estimates are biased and tend to strongly overestimate ν. A similar pattern is observed for the ML and REML estimates.

The above analysis demonstrated that the ML and REML approaches can be used to estimate the parameters required to calculate the elements of the covariance matrix. For estimating σ^2 it was shown that REML provided less biased results than the ML approach. This result was expected from theory. However, REML was not found to yield more accurate estimates of ϕ or ν. Indeed, in practice the REML estimates were found to be more variable and the numerical maximisation was less stable. The bias in the ML estimate of σ^2 (figure 7.31 (d)) can be partly compensated for by applying the adjustment given in equation 7.17, as recommended by Harris & Johnson (1996). Although this might be regarded as a less elegant and less accurate approach, it compensates for the bias in σ^2 *without* the other problems associated with REML.

If applying the above regression procedure on a small number of samples in an operational circumstance, the recommendation must be to apply all three approaches (ML, adjusted ML and REML). The user can then query each result in detail (for example, they can examine the shape of the likelihood surface to check the numerical maximisation). However, such detailed examination is not possible when working with several hundred simulations.

The regression and covariance model parameters were estimated using both the spherical and exponential models. The exponential model returned approximate values of the

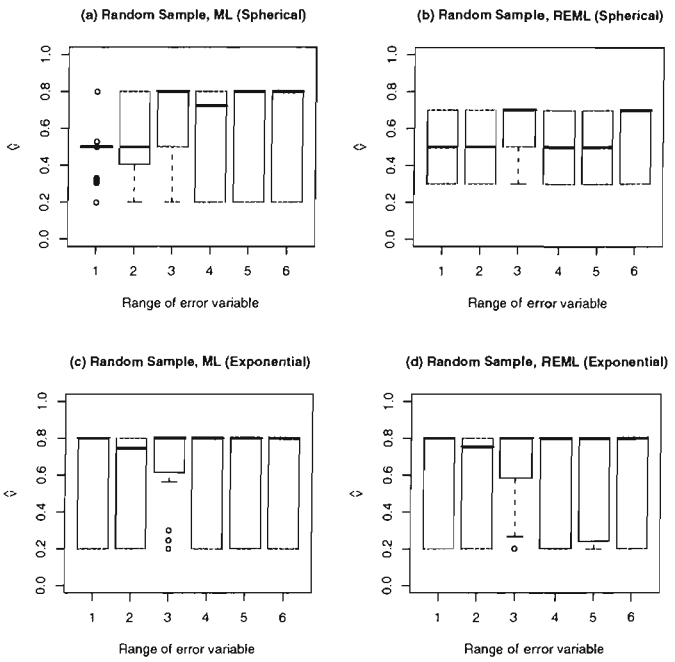


FIGURE 7.34: Boxplots showing the estimated value of ν for a given range of the error variable, ε. For the exponential model the *effective* range is shown. The results are shown for simple random sampling.

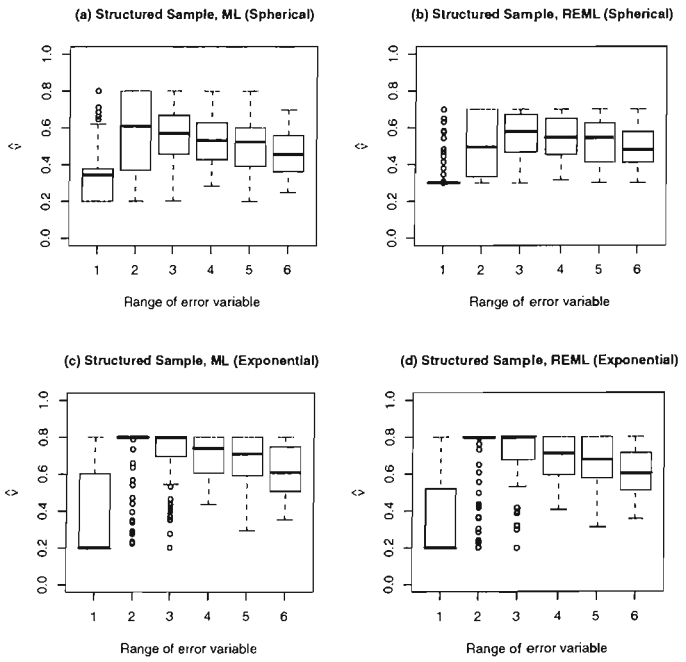


FIGURE 7.35: Boxplots showing the estimated value of ν for a given range of the error variable, ε. For the exponential model the *effective* range is shown. The results are shown for structured sampling on a 10×10 grid.

	ϕ	$\hat{\beta}_0$ OLS	ML	$\hat{\beta}_1$ OLS	ML
ME	1	-0.0007	-0.0005	-0.0014	-0.0009
	2	0.0025	0.0014	-0.0065	-0.0025
	3	-0.0244	-0.0296	-0.0076	-0.0045
	4	0.0089	0.0096	0.0166	0.0128
	5	0.0278	0.0168	0.0117	0.0071
	6	-0.0282	-0.0156	0.0119	0.0080
RMSE	1	0.0981	0.0979	0.1053	0.1052
	2	0.1265	0.0159	0.1027	0.1014
	3	0.1663	0.1652	0.1010	0.0894
	4	0.2086	0.2073	0.0976	0.0917
	5	0.2666	0.2550	0.1017	0.0890
	6	0.2987	0.2632	0.1050	0.0852

TABLE 7.13: Table showing the ME and RMSE for the estimates of $\beta = c(\beta_0, \beta_1)^T$ for the OLS and ML approaches.

estimated parameters, although these were less accurate than when the spherical model was applied. However, a choice between the spherical and exponential models could be made on the basis of the likelihood or the AIC.

Recall from equations 7.8 and 7.9 that $\hat{\beta}$ and $Cov(\hat{\beta})$ both depend on the covariance matrix, $\boldsymbol{\Sigma}$. The mean error (ME) and root mean square error (RMSE) for the estimates of $\beta = (\beta_0, \beta_1)^T$ are shown in table 7.13. This shows small reductions in the bias and an increase in precision of the point estimates of the model coefficients. Overall, the ML gives more *accurate* estimates of the model coefficients. This increase in accuracy is larger for larger values of the range, ϕ.

The standard errors of the coefficients require examination. These are obtained from the diagonal of $Cov(\hat{\beta})$. These are shown in figure 7.36 for the structured sample, where the parameters have been estimated using OLS and ML. For β_0 the standard error tended to increase whereas for β_1 it tended to decrease. This diagram provides a useful demonstration of the impact of $\boldsymbol{\Sigma}$ on $Cov(\hat{\beta})$. Such analysis is not presented in the literature, which tends to focus on the impact of the correlation structure on the estimate of σ^2 (e.g., Cook & Pocock 1983, Harris & Johnson 1996, Lark 2000). An increase (decrease) in the standard error leads to decrease (increase) in the associated t value. If this decrease is sufficiently large then it may lead to the situation where the null hypothesis that the coefficient is not significantly different to zero is *accepted.* In the example presented in this analysis there is no change in the outcome of this hypothesis test. However, it is important to realise that this may not always be the case and that there are examples where incorporating the correlation structure leads to fundamental changes in the model structure and conclusions drawn from an analysis (Ver Hoef et al. 2001, Lark & Cullis 2004, Lark & Webster 2006). Furthermore, the emergent effect on $Cov(\beta)$ of introducing the correlation structure is not straightforward to predict *a priori.* This is further illustrated when the

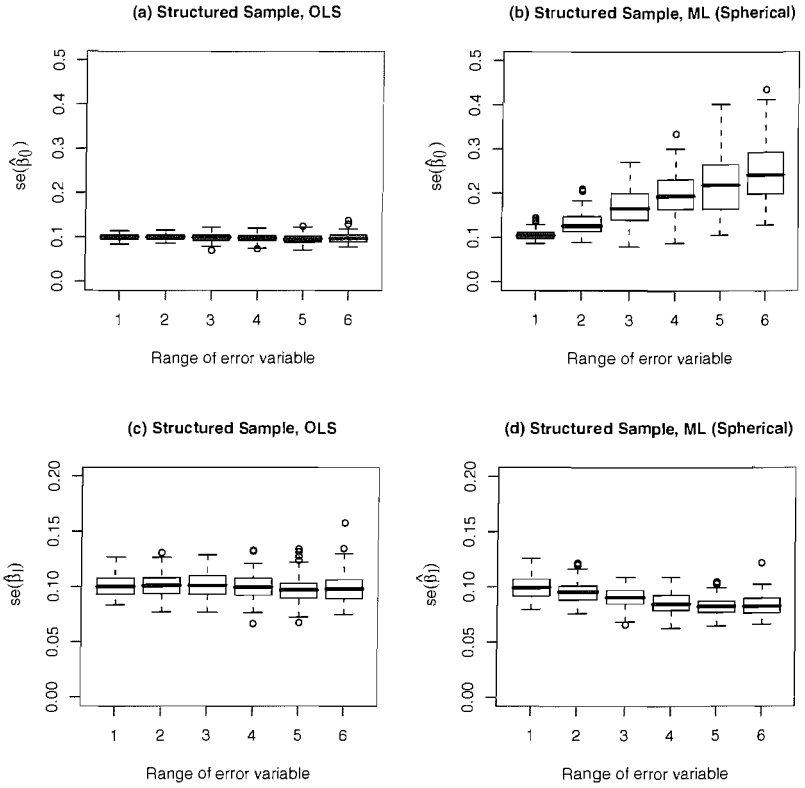


FIGURE 7.36: Boxplots showing the variability in the standard errors of the coefficients when estimated using OLS and ML for the structured sample. Figures (a) and (b) are for β_0 and (c) and (d) for β_1.

results for the real data are presented.

There are two final points that merit consideration. First, the above discussion noted that changes in $Cov(\beta)$ affect the associated t-values and can hence affect the results of hypothesis tests. Establishing the critical value for hypothesis testing requires specification of the degrees of freedom. However, where the residuals are correlated, it is not clear how many degrees of freedom there are. This point is acknowledged in the literature (Harris & Johnson 1996, Ver Hoef et al. 2001) but not fully addressed. Indeed, some authors simply calculate it in the same way as the uncorrelated case (Harris & Johnson 1996, Lark & Cullis 2004, Lark & Webster 2006). This issue will have a greater effect for small samples. This issue is not considered further in this thesis and is left as an area for future research. Second, the above analysis discussed parameter estimation but not prediction. This issue is given further consideration below.

In section 2.1.1 the difference between confidence and prediction intervals was discussed. The confidence interval gives an interval for $\hat{y}_0$, where $\hat{y}_0$ is a new point predicted using the regression line. The confidence interval is dependent on $Cov(\hat{\beta})$ and the t value associated with the specified p value (e.g., p=0.05 for the 95% confidence interval). Hence, the confidence interval quantifies the uncertainty in $\hat{y}_0$ due to uncertainty in the location of the regression line. The prediction interval also incorporates the residual variance, $\hat{\sigma}^2$.

Hence the prediction intervals will always be wider than the confidence intervals. Indeed, the data may show substantial scatter, leading to a large value for $\hat{\sigma}^2$ and wide prediction intervals, but, if the trend is clearly identifiable the confidence intervals may be narrow.

When there is no spatial autocorrelation in the residuals, prediction intervals are given for a single location (or block) (Neter et al. 1996). Hence, although the BLUP (best linear unbiased predictor) is $\hat{y}_0$, the true value of y_0 could lie anywhere within the prediction interval. However, for the case where there is spatial autocorrelation this approach makes no sense. This is because both the BLUP and a possible realisation will be affected by the neighbouring values of y and $\hat{y}$. Hence prediction and the prediction variance are as given in section 2.2.2.4 for kriging. These are recapped below:

$$y(\mathbf{u}_0) = x(\mathbf{u}_0)\beta + \boldsymbol{\Sigma}_0^T \boldsymbol{\Sigma}^{-1}(y(\mathbf{u}_0) - \mathbf{X}\beta) \tag{7.20}$$

where $x(\mathbf{u}_0)$ are the covariates at location $\mathbf{u}_0$, $\boldsymbol{\Sigma}$ is the covariance matrix for the residuals and, $\boldsymbol{\Sigma}_0 = (Cov[y(\mathbf{u}_0), y(\mathbf{u}_1)], \ldots, Cov(y(\mathbf{u}_0), y(\mathbf{u}_n)])^T$. Hence, the first term on the RHS is the prediction based solely on the regression model and the second term on the RHS modifies this value according to the values of the residuals at nearby data. The prediction error variance (kriging variance) is then given as:

$$\sigma^2(\mathbf{u}_0) = \sigma^2 - \boldsymbol{\Sigma}_0^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_0 + (x(\mathbf{u}_0) - \boldsymbol{\Sigma}_0^T \boldsymbol{\Sigma}^{-1}\mathbf{X})(\mathbf{X}^T \boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(x(\mathbf{u}_0) - \boldsymbol{\Sigma}_0^T \boldsymbol{\Sigma}^{-1}\mathbf{X})^T \tag{7.21}$$

which is the kriging variance. The RHS can be thought of as having three components. The first term is simply the residual variance (and variogram sill). The second term has a negative sign because, having measurements at neighbouring points *reduces* uncertainty in the prediction. However, the third term has a positive sign because uncertainty in the estimation of β acts to *increase* the uncertainty of prediction. If the measured values are very far from $\mathbf{u}_0$ (beyond the variogram range), then the second term on the RHS of both equations will approach zero.

If the objective is spatial prediction, equation 7.20 will give the BLUP image and equation 7.21 will give the prediction variance. This will apply both for interpolating spatially *between* measurements and for extrapolating spatially. In the latter case, the prediction will be dominated by $\hat{\beta}$ and will not be modified by neighbouring measurements. If a realisation from the surface is required, then this can be simulated using sequential simulation.

For many empirical models used in remote sensing the objective is to predict some biophysical property of the surface, such as biomass or green leaf area index (GLAI) (Curran & Williamson 1985, Dungan 1998). In such situations aspatial regression models or geostatistical models might be used to (i) model the relationship between field and remotely sensed data and (ii) predict the variable of interest across the image. In such cases, it is appropriate to predict the variable of interest using equations 7.20 and 7.21. Conditional simulation could then be used if it is necessary to simulate a "possible reality" for the

surface. However, in the case of the ELM, the situation is different. In this context, the ELM is a form of calibration and it is expected that the predicted image should resemble the input image. If conditional simulation was undertaken, this would yield an image that differed in spatial structure from the original image. Hence, realisations of the image should be limited to varying the regression line. This will require a modified version of equation 7.21. Additionally, this will be constrained by what is going on in neighbouring wavebands, as discussed in section 8.3.

7.5 Correlation and heteroskedasticity

Section 7.4 focused on estimating the regression parameters in the situation where there is spatial autocorrelation in the residuals. This assumed that σ^2 was uniform and, hence, the residuals were homoskedastic. This section relaxes the homoskedastic condition to deal with heteroskedastic residuals. This situation is likely to be widespread in reality and it was demonstrated in the case of the ELM in the first part of this chapter. However, to date, the problems of spatial autocorrelation and heteroskedasticity have been dealt with separately in both the GIS and remote sensing literature, although there are some examples in the epidemiological literature (e.g., Cook & Pocock 1983). Hence, the analysis presented in this section is novel in the context of remote sensing and GIS.

Under conditions of heteroskedasticity, the variance, σ_i^2 will vary across the image. Hence, as in equation 7.1, $\sigma_i^2 = \sigma^2/w_i$. Hence:

$$Cov(\varepsilon_i, \varepsilon_j) = \sigma_{ij} = \rho_{ij}\sigma_i\sigma_j\sqrt{1/w_i}\sqrt{1/w_j} \tag{7.22}$$

where ρ_{ij} is the correlation. Equation 7.16 can then be adapted to populate the elements of Σ:

$$\begin{aligned} \mathbf{\Sigma}_{ij} &= \sigma^2\sqrt{1/w_i}\sqrt{1/w_j}\nu f(\phi, \mathbf{h}_{ij}), i \neq j \\ &= \sigma^2, i = j \end{aligned} \tag{7.23}$$

The analysis proceeds first with simulated data where *known* weights are applied. This allowed testing and examination of the accuracy of the approach, without the potentially confounding effects of estimating the weights. Hence the covariance parameters (σ^2, ν and ϕ) were estimated using ML and REML but equation 7.23 was adopted rather than 7.16. As discussed in section 7.4 it is unusual for the weights to be known *a priori*. Hence, it is likely that the weights will need to be estimated using one of the procedures discussed in section 7.4. When the approach was applied to real data, estimated weights were used, since the true weights were unknown.

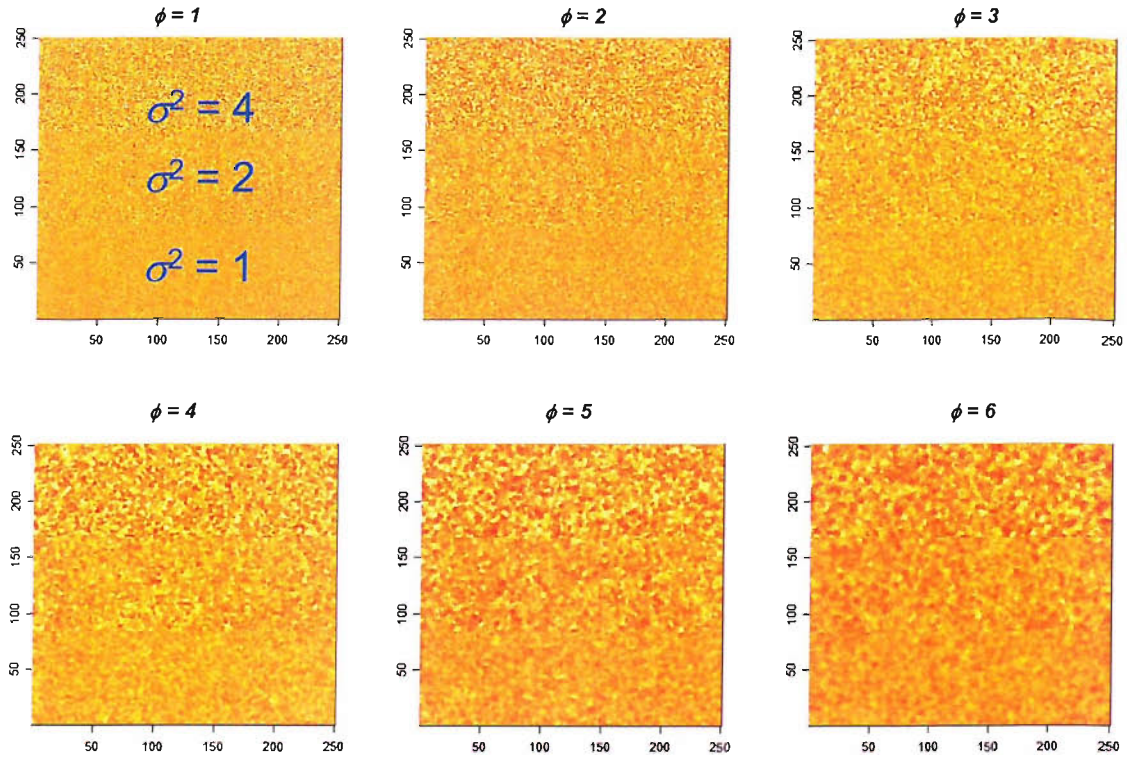


FIGURE 7.37: Unconditionally simulated images (one realisation) for the error term, ε. They have been modelled using a spherical model with $\sigma^2 = 1$, $\nu = 0.5$ and $\phi = 1, 2, \ldots, 6$.

7.5.1 Developing the approach for simulated data

For the case where σ^2 was allowed to vary across the image, the grid was divided into three strips, each 83 columns wide. Each strip in the image had an error variance, σ_i^2, of 1, 2 and 4 respectively. Hence, since $\sigma_i^2 = \sigma^2/w_i$, for $\sigma_i^2 = 1, 2, 4$, $w_i = 1, 0.5, 0.25$ and $\sigma^2 = 1$ as illustrated in figure 7.37. These six images were created using unconditional simulation for $\phi = 1, 2, \ldots, 6$. The image of $\mathbf{X}$ shown in figure 7.26 was used again and the image of $\mathbf{y}$ was again simulated using equation 7.19.

As before, 100 replicate samples were obtained (see section 7.4.1). However, this time the sample consisted of 300 data, with one 10 × 10 sample grid (or 100 random data points) lying in each strip. Hence, for each level of σ_i^2, the sample size was the same as that used previously. However, since there were three levels of σ_i^2, the overall sample size was larger.

Having obtained the simulated images of $\mathbf{y}$ and $\mathbf{X}$ parameter estimation was performed on each sample using WLS, ML and REML. This was repeated 100 times to establish the accuracy of the parameter estimation. Note that, in this case, WLS was used instead of OLS.

7.5.1.1 Results

Results in this section are presented in the same format as section 7.4.1.1.

Figure 7.38 shows the estimated value of σ for WLS, ML and REML when the spherical model is adopted. This showed that, under WLS, estimates of σ^2 were biased. However, both ML and REML led to unbiased estimators of σ. Unlike the case presented in section 7.4.1.1 ML does not provide more accurate estimates than REML. However, it should be recalled that the sample comprised of three sets of 100 data points (one for each "strip"), giving a total sample size of 300. The analysis presented in section 7.4.1.1 was based on a sample size of 100 for the entire 249×249 image. Hence, it is important not to confound the effect of increasing the sample size with heteroskedasticity. Indeed, when the sample size was increased to 3×100, but with $\sigma_i^2 = 1$ for all three strips both ML and REML were found to give unbiased estimates of σ^2. Hence this is attributed to the sample size, rather than the presence of heteroskedasticity.

Figure 7.39 shows the estimated value of ϕ. It is interesting to note that under random sampling both ML and REML yield approximate estimates for ϕ. The same pattern was observed under homoskedastic conditions for a sample size of 3×100. This was an important result since it implied that the larger sample size provided a sufficiently large sample of nearby data to allow an approximate estimate of ϕ. Importantly, this also showed that the model residuals were correlated and the fact that random sampling was employed does not mean that independence can be assumed. For ML and REML, the estimates of ϕ were unbiased and precise by comparison to figures 7.32 and 7.33

The estimates of ν for the random and structured sample and the spherical model are shown in figure 7.40. As for σ^2 and ϕ, the large sample size yielded accurate estimates of ν for the structured sample, particularly for $\phi \geq 3$. For the random sample the estimates were highly inaccurate for $\phi \leq 3$. For $\phi > 3$, the estimates, although still highly variable, were more accurate than for those for the smaller sample presented in figure 7.34.

The analysis presented above reveals two clear conclusions:

1. The larger sample size allows more accurate estimation of all three parameters for the structured sample. There is no substantial difference between the estimates obtained using ML and REML, a result also found by Diggle et al. (2003). Importantly, the large sample size also allows "detection" of the spatial correlation in the random sample. This is most clearly shown for $\hat{\phi}$ (figure 7.39), but can also be observed for $\hat{\nu}$ (figure 7.40). This is an important result since, even if a random sample is used, independence cannot be assumed (see also Lark 2000). Hence it is necessary to test whether the residuals are correlated.

2. These results show that the extension to the heteroskedastic case, given in equation 7.23, allowed accurate estimation of the covariance parameters. On a practical note, the algorithm was found to yield stable results. However, in this analysis, the weights used were those that were known *a priori*. In reality, the weights would need to be estimated, as discussed previously. This analysis was not conducted for the simulated images, but is presented for real data in section 7.6.1.

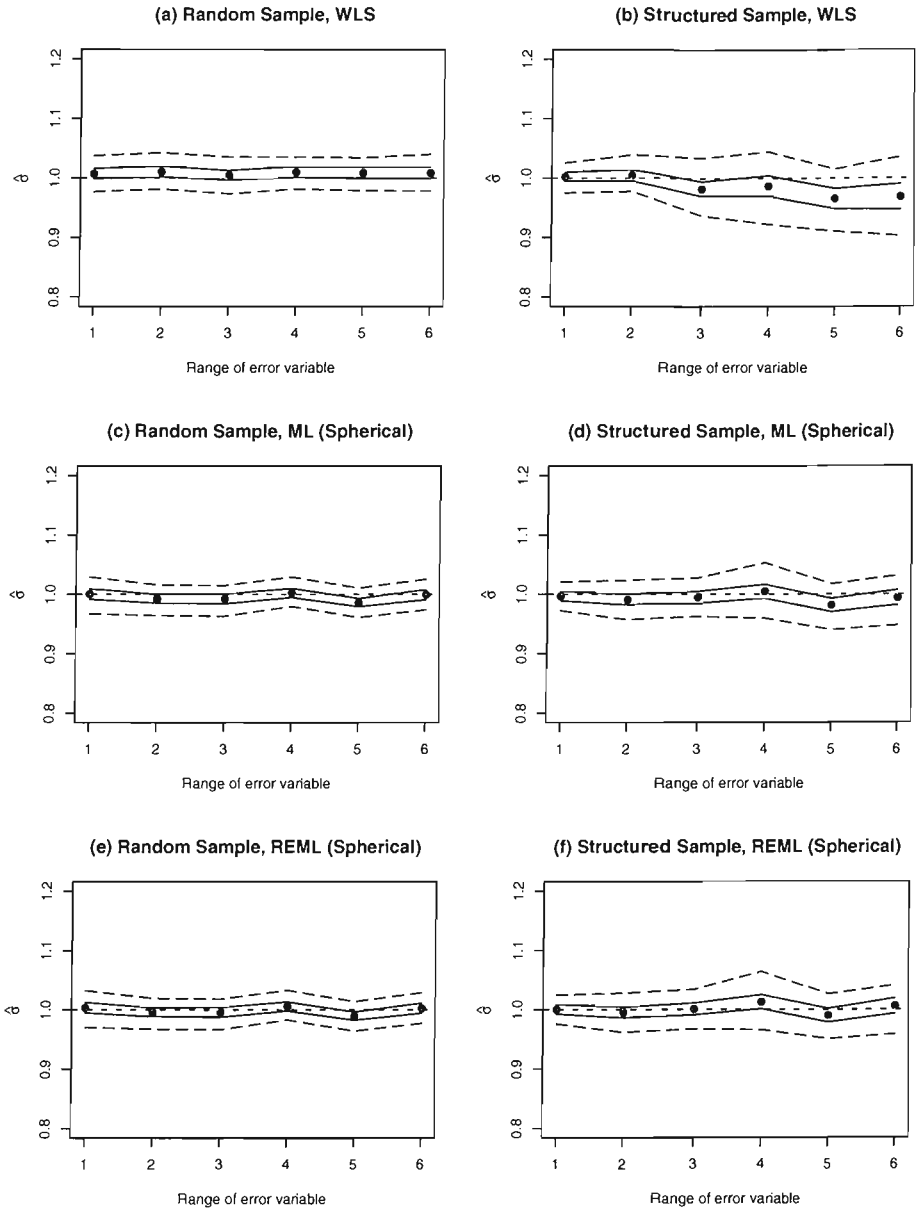


FIGURE 7.38: Estimated value of σ of a given range of the error variable. The points show the mean estimate of σ. Solid lines show the 95% confidence intervals on the mean and the dashed line show the upper and lower quartiles.

Finally, as in section 7.4.1.1, adoption of ML or REML led to an increase in the standard error of β_0 but a decrease in the standard error of β_1.

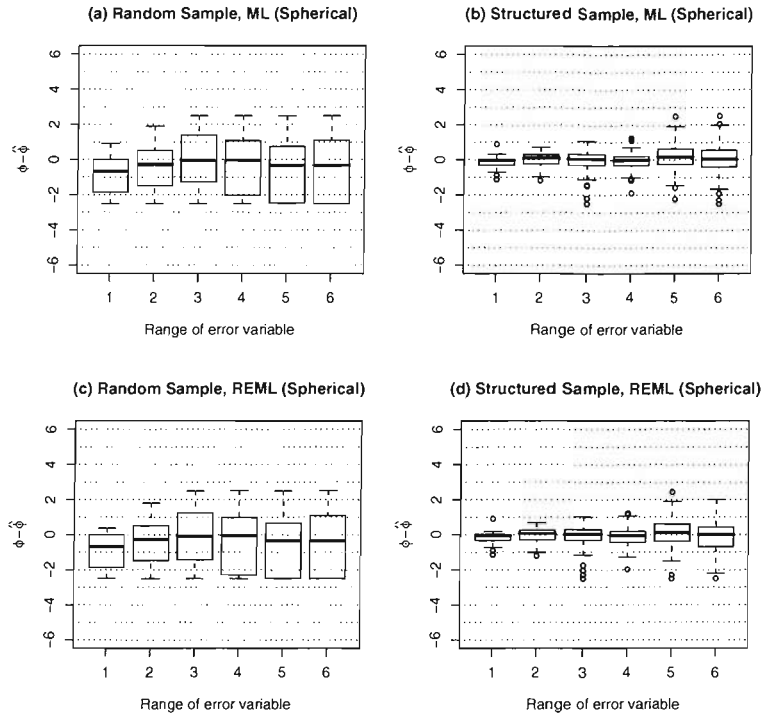


FIGURE 7.39: Boxplots showing $\phi - \hat{\phi}$ for a given range of the error variable, ε. The results are shown for the case where the spherical model is adopted for both random and structured sampling.

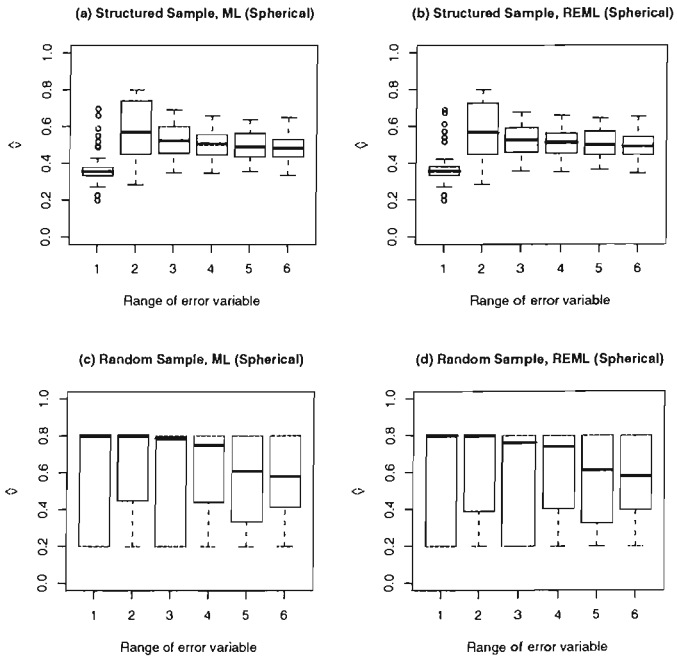


FIGURE 7.40: Boxplots showing the estimated value of ν for a given range of the error variable, ε. Figures a and b show the ML and REML results for the structured sample ($3 \times 10 \times 10$ square grid) whereas c and d show the results for the random sample (3×100).

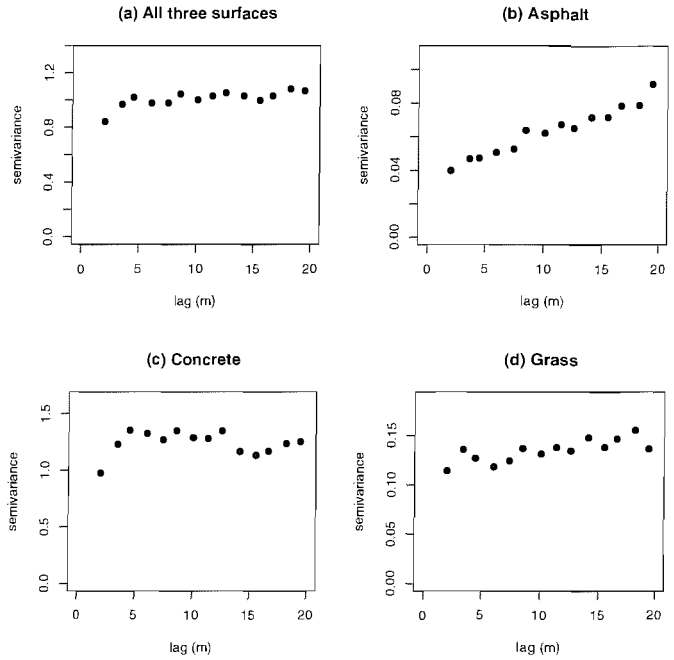


FIGURE 7.41: Plots showing the residuals from the OLS regression model for MMR band 1 regressed on ATM band 2. The MMR data were aggregated using the simple averaging approach. The variograms for all three surfaces (a) were calculated for the standardised residuals. The variograms for the individual GTs were calculated using the non-standardised residuals.

7.6 Applying the approach for real data

Having investigated the ML and REML approaches to regression with correlated residuals, the approach was applied to real data – the data used for parameter estimation and implementation of the ELM. Throughout this section, the **y** (reflectance) data used were those aggregated using the simple averaging approach and all pixels containing one or more field measurements were retained. The aggregation was performed using the simple averaging approach, owing to the previously discussed limitations of using kriged or conditionally simulated blocks.

Figures 7.41 to 7.44 show the variograms of the OLS residuals for the four MMR bands. The variogram of the standardised residuals for all surfaces are also shown. These variograms show clear evidence of spatial structure for the asphalt and concrete surfaces in all three wavebands. It is interesting to note that the form of this spatial structure varies between bands 1 to 3 (visible wavebands) and band 4 (near-infrared waveband). Similar results were observed by Atkinson & Emery (1999), who found that the spatial structure of the reflectance of vegetation varies between visible and near-infrared wavelengths. Further evidence for spatial structure in the residuals was found when a variogram was estimated for the OLS residuals using the ML approach. Comparison of the variogram and non-spatial model showed strong evidence for spatial structure for all targets and all wavebands, although the evidence is weaker for the grass target in visible wavebands.

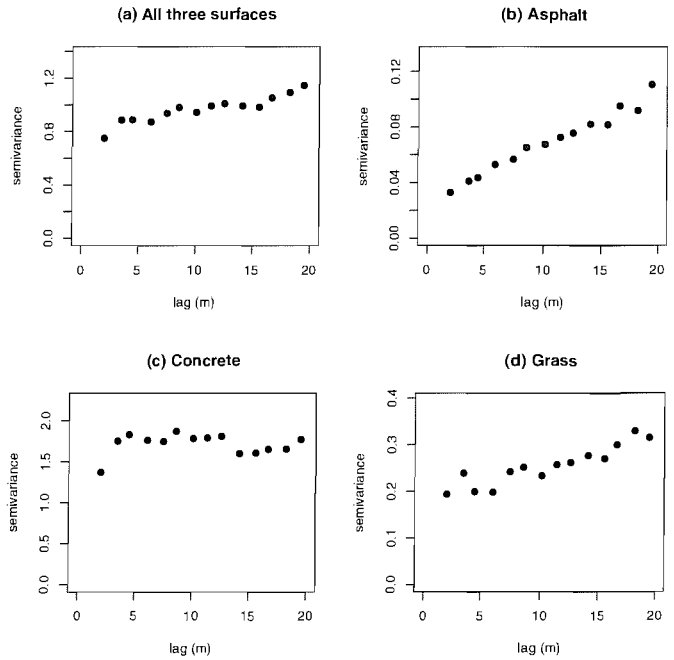


FIGURE 7.42: Plots showing the residuals from the OLS regression model for MMR band 2 regressed on ATM band 3. Other details are as for figure 7.41

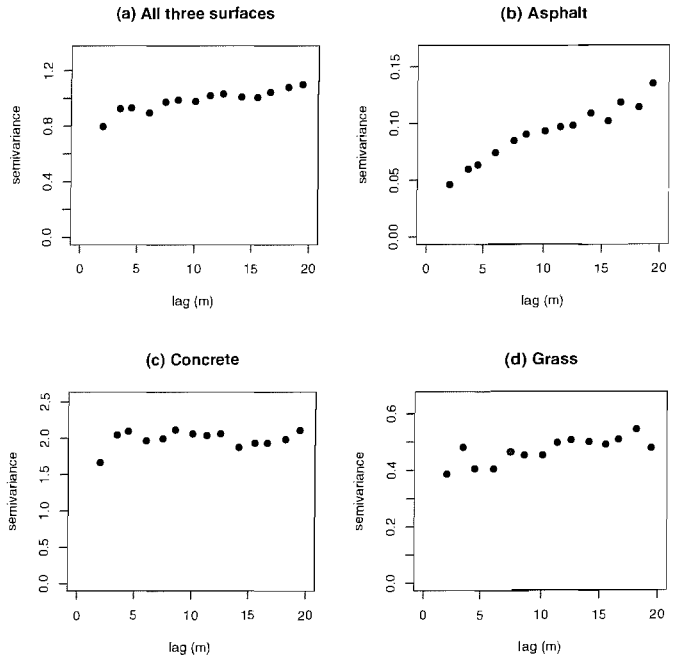


FIGURE 7.43: Plots showing the residuals from the OLS regression model for MMR band 3 regressed on ATM band 5. Other details are as for figure 7.41

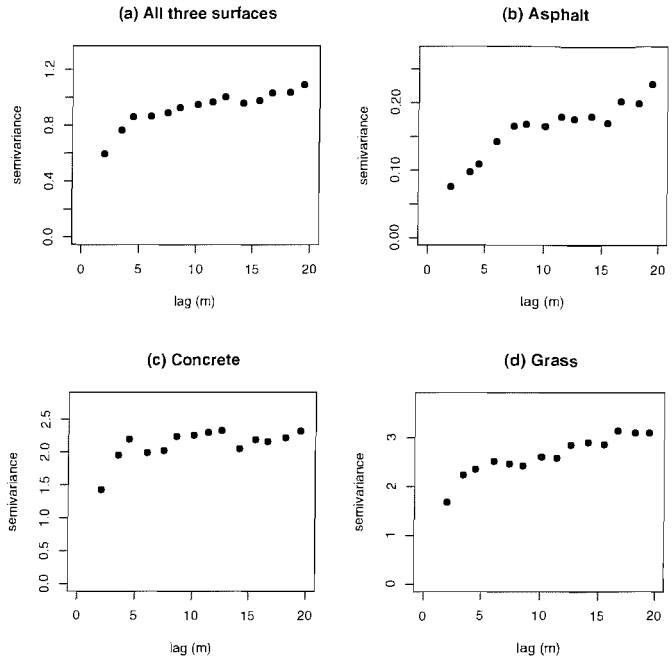


FIGURE 7.44: Plots showing the residuals from the OLS regression model for MMR band 4 regressed on ATM band 7. Other details are as for figure 7.41

The fact that there is evidence for spatial structure in the residuals means that this should be accounted for in the regression model. However, implementing the spatial regression model does present a challenge. It will be recalled that the covariance matrix, $\mathbf{\Sigma}$ had three parameters: σ^2, ν and ϕ. However, the shape and magnitude of the variogram varies between targets. The model presented in previous sections used a single model, although it accounted for variation in σ^2 through the weighting structure. The estimation of appropriate weights was discussed in detail in section 7.1. However, the ML and REML models do not allow provision for varying the form of the variogram, given by ϕ and ν. Hence, the model will only give an approximate and general estimation of the correlation structure.

It is not clear how the ML model can be adapted to allow the form of the variogram to vary across the image. Within the regression kriging framework, discussed in section 2.2.2.4, this could be addressed in an *ad hoc* manner. The residuals from the OLS or WLS model could be used to estimate the variogram parameters which could then be used to construct the correlation matrix. However, such an approach would not have the optimal properties of the ML approach (Hengl et al. 2004, Lark & Cullis 2004, Lark & Webster 2006). Furthermore, although this approach would account for the effect of correlation on estimation, it is less useful for prediction since the interpolation component could not be applied for surfaces that do not lie within the training data set used to build the models. It might be argued that, since the ELM is a form of calibration, primary interest should lie in accurate estimation of β and its associated confidence intervals (as discussed previously). However, for more general purposes the prediction interval will be required. Indeed, this ability to use the model at locations not included in the training set

	MMR B1 on ATM B2	MMR B2 on ATM B3	MMR B3 on ATM B5	MMR B4 on ATM B7
$\hat{\beta}_0$	-2.483	-0.6142	-1.8526	-3.8777
$s.e.(\hat{\beta}_0)$	0.0987	0.1208	0.1212	0.2778
p	< 0.01	< 0.01	< 0.01	< 0.01
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\beta}_1$	0.3021	0.2949	0.3529	0.5415
$s.e.(\hat{\beta}_1)$	0.0024	0.0025	0.0025	0.0048
p	< 0.01	< 0.01	< 0.01	< 0.01
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\sigma}^2$	0.6121	0.9721	1.1128	2.262
$\hat{\sigma}$	0.7824	0.9859	1.055	1.504
R^2	0.9738	0.9709	0.9785	0.9671

TABLE 7.14: Table showing the result of parameter estimation of the ELM regression model using the simple averaging approach and OLS. All pixels containing one or more field measurement are included.

is one of the advantages of regression over cokriging (Dungan 1998, Lark 2000). Despite its disadvantage, this *ad hoc* RK approach is given some consideration in the results section.

7.6.1 Results

In order to evaluate the performance of the combined correlated and heteroskedastic residuals the results of three model implementations for the simple averaging approach are evaluated. These include the OLS case, presented in table 5.10, section 5.1 and is replicated below (table 7.14). Next is the WLS case, where the weights are computed as the inverse of the variance of the OLS residuals for each GT (see section 7.2.2). The WLS results are presented in table 7.15. Finally the results of the combined correlated and heteroskedastic residuals model (WML) are presented in table 7.17. For the WML model, the model was fitted using both the ML and REML approach and the results compared. Both the spherical and exponential correlation models were considered and evaluated.

The difference between the OLS and WLS parameter estimates varied depending on the waveband examined. For MMR band 1 and MMR band 3 the difference in the estimates of the model coefficients and their associated standard errors is small. Explanation for this can be found in figure 5.13, which shows that the OLS regression line passes through the centre of the cloud of points associated with each GT. Hence, when differential weights are applied to each GT, substantial change in the location of the regression line would not be expected. However, the confidence intervals on the coefficients and on the regression line may change. For MMR band 2, the situation was more complex; in this case the GTs do not lie on a straight line, and applying different weightings to each GT caused a small but substantial change in the estimates of the regression coefficients. Furthermore,

	MMR B1 on ATM B2	MMR B2 on ATM B3	MMR B3 on ATM B5	MMR B4 on ATM B7
$\hat{\beta}_0$	-2.5476	-0.487	-1.8468	-3.5301
$s.e.(\hat{\beta}_0)$	0.0653	0.1089	0.0978	0.1272
p	< 0.01	< 0.01	< 0.01	< 0.01
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\beta}_1$	0.3046	0.2835	0.351	0.5345
$s.e.(\hat{\beta}_1)$	0.0020	0.0030	0.0027	0.0032
p	< 0.01	< 0.01	< 0.01	< 0.01
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\sigma}^2$	1.009	1.4256	1.0016	1.093
$\hat{\sigma}$	1.0173	1.1941	1.001	1.045
R^2	0.9812	0.9541	0.9743	0.9849
$w_{asphalt}$	11.834	10.0761	8.1795	4.8504
$w_{concrete}$	0.8288	0.5911	0.5022	0.4522
w_{grass}	6.9545	3.5011	1.9885	0.3329
$w^{relative}_{asphalt}$	1.0000	1.0000	1.0000	1.0000
$w^{relative}_{concrete}$	0.0700	0.0587	0.0614	0.0932
$w^{relative}_{grass}$	0.5877	0.3475	0.2431	0.0686

TABLE 7.15: Table showing the result of parameter estimation of the ELM regression model using the simple averaging approach and WLS. All pixels containing one or more field measurement are included.

the standard error on β_1 was widened, whilst the standard error on β_0 was reduced. The effect of these changes is illustrated in table 7.16. This shows that the location of the regression line and the confidence intervals around that regression line differ for the OLS and WLS case – although the impact of that change is greatest for large radiance values. Finally, for MMR band 4 small changes were observed in the estimates of β_0 and β_1. Their associated standard errors were also narrower for the WLS approach. Note that the ordering of the reflectance values between the targets differs for the visible and near-infrared bands. For the NIR, grass is the brightest GT, whereas it is the darkest target in the visible bands. It also shows larger variance, relative to the other targets in the NIR. As a result, the asphalt target, which shows the lowest variance (and the highest weight) in all three bands, is not bracketed by grass and concrete. Hence, when the weighting is applied, it is more likely to exert control over the regression parameter estimates.

Notwithstanding any change in the parameter estimates, applying the weighting scheme will lead to a change in the value of σ_i^2 associated with each target. In all cases, this is important for prediction.

The results from the combined heteroskedastic and correlated residual model are shown in table 7.17. For MMR band 4 the spherical model was chosen, since the exponential model fails to converge numerically. The spherical model was also a reasonable choice based on the variograms presented in figure 7.44. Both ML and REML yield identical results to two

Radiance	$\hat{y}_{OLS}$	95% CI	$\hat{y}_{WLS}$	95% CI	$\hat{y}_{WLS} - \hat{y}_{OLS}$	Example Landcover
0	-0.61	0.45	-0.49	0.43	0.13	-
10	0.234	0.37	2.35	0.32	0.01	Dense vegetation
20	5.28	0.29	5.19	0.21	-0.1	Grass
30	8.23	0.22	8.02	0.13	-0.22	Grass
40	11.18	0.18	10.85	0.13	-0.33	Urban surfaces (e.g., asphalt)
50	14.13	0.18	13.69	0.21	-0.44	Urban surfaces
60	17.08	0.22	16.52	0.31	-0.56	Urban surfaces
70	20.03	0.29	19.36	0.42	-0.67	Concrete
80	22.98	0.36	22.19	0.54	-0.79	Concrete
90	25.93	0.44	25.03	0.65	-0.9	Sand
100	28.88	0.53	27.86	0.77	-1.02	Metallic surfaces (e.g., sheds)

TABLE 7.16: Table showing the predicted values of reflectance for MMR band 2 (regressed on ATM band 3) for a range of radiance values. Results are shown for the OLS and WLS regression models (using the simple averaging blocks). The 95% confidence intervals on the regression line are also shown.

decimal places, which was judged to be reasonable given that the likelihood is maximised numerically. The estimate of σ^2 was slightly larger than that obtained using WLS. This was expected given that, as discussed previously, failure to account for correlation tends to lead to an underestimate of σ^2. Furthermore, the correlation structure leads to an increase in the standard errors on the estimates of the regression coefficients relative to the WLS case. Hence, incorporating the correlation structure has important implications for quantifying the uncertainty in estimation and prediction. This is illustrated in table 7.18. These are important results, since they show that modelling the correlation leads to an increase in the confidence interval on the regression line, particularly for large radiance values. Importantly, for the NIR waveband, these correspond to vegetated surfaces which are often of considerable interest to users of remotely sensed data.

Finally, the results for the visible wavebands (MMR bands 1, 2 and 3) are now considered. Both the spherical and exponential models estimated very long range variation for both the ML and REML approaches. Associated with this were estimates of σ^2 that seem unrealistically high by comparison to the WLS approach. The simulation experiments do not suggest that the increase should be so large. Hence, these results are attributed to the inability of the combined heteroskedastic and correlated regression model to cope with the complicated spatial structure shown in Figures 7.41, 7.42 and 7.43. One interpretation might be that, for the visible wavebands, the spatial correlation structure is substantially different between targets and cannot be modelled using a single value for ν and ϕ. For the NIR waveband the spatial structure appears (visually) to be much more similar between GTs. Furthermore, the long range variation in the visible wavebands, particularly for the asphalt surface, may be indicative of some unexplained variation (Webster & Oliver 2001, Ver Hoef et al. 2001) that might be attributed to a trend (on location) or another covariate. However, incorporating a trend on location is impractical given that a regression model based on attribute (radiance) is required. The possibility of including further covariates

	MMR B1 on ATM B2	MMR B2 on ATM B3	MMR B3 on ATM B5	MMR B4 on ATM B7
$\hat{\beta}_0$	-0.5717	0.7191	-0.4813	-3.377
$s.e.(\hat{\beta}_0)$	0.5054	0.5702	0.5784	0.1981
p	> 0.1	< 0.1	> 0.1	< 0.01
$p < 0.05$	No	No	No	Yes
$p < 0.01$	No	No	No	Yes
$\hat{\beta}_1$	0.2421	0.246	0.3079	0.5294
$s.e.(\hat{\beta}_1)$	0.0134	0.0147	0.0155	0.0052
p	< 0.01	< 0.01	< 0.01	< 0.01
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
Correlation model	Spherical	Spherical	Spherical	Spherical
$\hat{\sigma}^2$	2.2622	1.7662	1.5304	1.107
$\hat{\sigma}$	1.5041	1.329	1.2371	1.0521
$\hat{\nu}$	0.6674	0.6274	0.5055	0.8146
$\hat{\phi}$	179.9448	95.2479	96.6057	6.1848
R^2	0.6377	0.7039	0.7584	0.9261
$w_{asphalt}$	11.834	10.0761	8.1795	4.8504
$w_{concrete}$	0.8288	0.5911	0.5022	0.4522
w_{grass}	6.9545	3.5011	1.9885	0.3329
$w_{asphalt}^{relative}$	1.0000	1.0000	1.0000	1.0000
$w_{concrete}^{relative}$	0.0700	0.0587	0.0614	0.0932
$w_{grass}^{relative}$	0.5877	0.3475	0.2431	0.0686

TABLE 7.17: Table showing the result of parameter estimation of the ELM regression model using the simple averaging approach and the combined heteroskedastic and correlated residuals (WML) approach. All pixels containing one or more field measurement are included. The parameters were estimated using the maximum likelihood approach with a variogram embedded to account for the spatial correlation.

was not explored, although this subject is discussed in chapter 8.

In order to address the correlation in the residuals the *ad hoc* RK type approach was adopted. This is presented in table 7.19. The correlation structure was estimated from the OLS residuals for each GT using the MBG-ML approach. This was used to populate the correlation matrix which was then "plugged-in" to obtain the GLS estimator of β. The resulting parameter estimates are presented in table 7.19. For the visible wavebands the variogram range is large for the asphalt surface, but shorter for the concrete and grass surfaces. The predicted values of y (y_{RK}) together with the associated confidence intervals are presented in tables 7.20 to 7.23. For bands 1 and 3 $\hat{y}_{RK}$ (table 7.20 and 7.22) is similar to $\hat{y}_{OLS}$ and $\hat{y}_{WLS}$, although the confidence intervals are wider. For band 2 (table 7.21), the results are interesting, since $\hat{y}$ obtained for RK is further modified relative to the OLS and WLS cases. The confidence intervals are also wider. This demonstrates that failure to account for spatial correlation can affect both the *predicted value* and the *uncertainty estimate*. Finally, the RK predictions and confidence intervals for band 4 (table 7.23) were similar to those obtained for the WML approach.

Radiance	$\hat{y}_{OLS}$	CI	$\hat{y}_{WLS}$	CI	$\hat{y}_{WML}$	CI	Example Landcover
0	-3.89	1.08	-3.53	0.50	-3.38	0.77	-
10	1.53	0.90	1.81	0.38	1.92	0.58	Water
20	6.94	0.73	7.16	0.27	7.21	0.41	Shallow water
30	12.35	0.56	12.50	0.19	12.50	0.27	Urban surfaces
40	17.77	0.41	17.85	0.16	17.80	0.24	Urban surfaces
50	23.18	0.30	23.19	0.22	23.09	0.36	Concrete
60	28.59	0.29	28.54	0.32	28.39	0.52	Concrete
70	34.01	0.38	33.88	0.43	33.68	0.71	Grass
80	39.42	0.53	39.23	0.55	38.98	0.90	Grass
90	44.84	0.69	44.57	0.67	44.27	1.10	Dense vegetation
100	50.25	0.87	49.92	0.79	49.57	1.30	Dense vegetation

TABLE 7.18: Table showing the predicted values of reflectance for MMR band 4 (regressed on ATM band 7) for a range of radiance values. Results are shown for the OLS, WLS and combined heteroskedastic and correlated residuals (WML) regression models (using the simple–averaged blocks). The 95% confidence intervals on the regression line are also shown.

	MMR B1 on ATM B2	MMR B2 on ATM B3	MMR B3 on ATM B5	MMR B4 on ATM B7
$\hat{\beta}_0$	-2.4748	-0.7205	-1.909	-3.2242
$s.e.(\hat{\beta}_0)$	0.0756	0.2275	0.1652	0.2286
p	< 0.01	< 0.01	< 0.01	< 0.01
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\beta}_1$	0.2996	0.2949	0.3527	0.5249
$s.e.(\hat{\beta}_1)$	0.029	0.0044	0.0034	0.0052
p	< 0.01	< 0.01	< 0.01	< 0.01
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\sigma}^2$	1.013	1.0549	1.0104	1.027
$\hat{\sigma}$	1.0065	1.0271	1.0052	1.0134
Correlation model	Spherical	Spherical	Spherical	Spherical
$\hat{\phi}_{asphalt}$	48.12	54.11	40.38	11.41
$\hat{\nu}_{asphalt}$	0.5828	0.7195	0.6214	0.6941
Correlation model	Spherical	Spherical	None	Spherical
$\hat{\phi}_{concrete}$	3.418	3.472	-	4.213
$\hat{\nu}_{concrete}$	1	1	0	1
Correlation model	None	Spherical	Spherical	Exponential
$\hat{\phi}_{grass}$	-	29.81	16.44	7.333
$\hat{\nu}_{grass}$	0	0.333	0.1657	0.712
R^2	0.9676	0.9816	0.9865	0.9109
$w_{asphalt}$	11.834	10.0761	8.1795	4.8504
$w_{concrete}$	0.8288	0.5911	0.5022	0.4522
w_{grass}	6.9545	3.5011	1.9885	0.3329
$w^{relative}_{asphalt}$	1.0000	1.0000	1.0000	1.0000
$w^{relative}_{concrete}$	0.0700	0.0587	0.0614	0.0932
$w^{relative}_{grass}$	0.5877	0.3475	0.2431	0.0686

TABLE 7.19: Table showing the result of parameter estimation of the ELM regression model using the simple averaging approach and the combined heteroskedastic and correlated residuals RK approach. All pixels containing one or more field measurement are included. The parameters were estimated using the regression kriging (RK) type approach where the spatial correlation structure is estimated from the OLS residuals for each GT.

Radiance	$\hat{y}_{OLS}$	CI	$\hat{y}_{WLS}$	CI	$\hat{y}_{RK}$	CI	Example Landcover
0	-2.46	0.36	-2.55	0.26	-2.47	0.30	-
10	0.55	0.29	0.50	0.18	0.52	0.20	Dense vegetation
20	3.56	0.21	3.54	0.12	3.52	0.12	Grass
30	6.57	0.16	6.59	0.08	6.51	0.13	Grass
40	9.58	0.14	9.63	0.11	9.51	0.22	Urban surfaces (e.g., asphalt)
50	12.59	0.17	12.68	0.17	12.50	0.32	Urban surfaces
60	15.60	0.24	15.73	0.25	15.50	0.43	Concrete
70	18.61	0.31	18.77	0.32	18.50	0.54	Concrete
80	21.62	0.39	21.82	0.40	21.49	0.66	Metallic surfaces (e.g., sheds)
90	24.63	0.48	24.86	0.48	24.49	0.77	Metallic surfaces (e.g., sheds)
100	27.64	0.56	27.91	0.56	27.48	0.89	Sand

TABLE 7.20: Table showing the predicted values of reflectance for MMR band 1 (regressed on ATM band 2) for a range of radiance values. Results are shown for the OLS, WLS and the RK type regression model. Field measurements are aggregated to pixel-sized supports using the simple averaging approach where every pixel containing one or more field measurement is retained. The 95% confidence intervals on the regression line are also shown.

Radiance	$\hat{y}_{OLS}$	CI	$\hat{y}_{WLS}$	CI	$\hat{y}_{RK}$	CI	Example Landcover
0	-0.61	0.45	-0.49	0.43	-0.72	0.89	-
10	2.32	0.37	2.35	0.32	2.23	0.74	Dense vegetation
20	5.26	0.29	5.18	0.21	5.18	0.59	Grass
30	8.20	0.22	8.02	0.13	8.13	0.46	Grass
40	11.14	0.18	10.85	0.13	11.08	0.36	Urban surfaces (e.g., asphalt)
50	14.07	0.18	13.69	0.21	14.03	0.34	Urban surfaces
60	17.01	0.22	16.52	0.31	16.97	0.40	Urban surfaces
70	19.95	0.28	19.36	0.43	19.92	0.51	Concrete
80	22.89	0.36	22.19	0.54	22.87	0.65	Concrete
90	25.82	0.44	25.03	0.65	25.82	0.80	Sand
100	28.76	0.53	27.86	0.77	28.77	0.96	Metallic surfaces (e.g., sheds)

TABLE 7.21: Table showing the predicted values of reflectance for MMR band 2 (regressed on ATM band 3) for a range of radiance values. Details are as given in table 7.20.

Radiance	$\hat{y}_{OLS}$	CI	$\hat{y}_{WLS}$	CI	$\hat{y}_{RK}$	CI	Example Landcover
0	-1.82	0.45	-1.85	0.38	-1.91	0.65	-
10	1.69	0.36	1.66	0.28	1.62	0.53	Dense vegetation
20	5.21	0.29	5.17	0.19	5.14	0.42	Grass
30	8.73	0.23	8.68	0.12	8.67	0.33	Asphalt
40	12.24	0.19	12.19	0.13	12.20	0.28	Urban surfaces (e.g., asphalt)
50	15.76	0.20	15.70	0.20	15.73	0.29	Urban surfaces
60	19.28	0.24	19.21	0.30	19.25	0.35	Concrete
70	22.79	0.31	22.72	0.40	22.78	0.44	Concrete
80	26.31	0.38	26.23	0.50	26.31	0.55	Sand
90	29.83	0.47	29.74	0.61	29.83	0.67	Metallic surfaces (e.g., sheds)
100	33.35	0.55	33.25	0.72	33.36	0.79	Metallic surfaces (e.g., sheds)

TABLE 7.22: Table showing the predicted values of reflectance for MMR band 3 (regressed on ATM band 5) for a range of radiance values. Details are as given in table 7.20.

Radiance	$\hat{y}_{OLS}$	CI	$\hat{y}_{WLS}$	CI	$\hat{y}_{RK}$	CI	Example Landcover
0	-3.89	1.08	-3.53	0.50	-3.22	0.90	-
10	1.53	0.90	1.81	0.38	2.02	0.71	Water
20	6.94	0.73	7.16	0.27	7.27	0.53	Shallow water
30	12.35	0.56	12.50	0.19	12.52	0.38	Urban surfaces
40	17.77	0.41	17.85	0.16	17.77	0.30	Urban surfaces
50	23.18	0.30	23.19	0.22	23.02	0.34	Concrete
60	28.59	0.29	28.54	0.32	28.27	0.48	Concrete
70	34.01	0.38	33.88	0.43	33.52	0.65	Grass
80	39.42	0.53	39.23	0.55	38.77	0.84	Grass
90	44.84	0.69	44.57	0.67	44.01	1.03	Dense vegetation
100	50.25	0.87	49.92	0.79	49.26	1.23	Dense vegetation

TABLE 7.23: Table showing the predicted values of reflectance for MMR band 4 (regressed on ATM band 7) for a range of radiance values. Details are as given in table 7.20.

7.6.2 Additional results

The results presented in section 7.6.1 utilised simple-averaged aggregations of reflectance for each pixel. Importantly, all pixels containing *one or more* field measurements were included in the analysis. However, the decision might be made to include only those pixel which contain more than a certain number of field measurements – a point that is given further discussion in section 8.1. Clearly, for the data set used in this thesis, it is not possible to set a threshold higher than two.

In section 5.4.1, OLS results were presented where only pixels containing *two or more* field measurements were included in the analysis. The decision to include pixels containing one or more (denoted G0) or two or more (denoted G1) pixels made only a small difference to the estimates of the OLS coefficients. However, the estimates of σ^2 were lower for the G1 case, although the confidence intervals on the regression line tended to be wider. This increased width in the confidence interval arose because this is dependent on $Cov(\beta)$, rather than just σ^2. It was suggested that it would be preferable to use the G1 data set in preference to G0 since, by averaging over a greater number of field measurements a more accurate prediction of the per-pixel value is obtained (Curran & Williamson 1985). However, it was also noted that there are only 111 G1 pixels whereas there are 436 G0 pixels. For the analysis undertaken in the current chapter the G0 pixels were used, in order to give a larger number of pixels for the geostatistical analysis. Nevertheless, it is important to give some consideration to the likely impact of choosing G1.

The results for G1 are discussed briefly here. First, the relative weights, obtained from the variance of the OLS residuals were quite different for G1 *vs.* G0. This is clear if tables 7.17 and 7.24 are compared. In particular, the relative weight of concrete is lower for the G1 case. It is proposed that this difference is due to the fact that the spatial structure differs between GTs and specifically due to the relative magnitude of the *within pixel* and *between pixel* variability. The different weighting for G0 and G1 has consequences for the confidence intervals on the regression line and for prediction intervals.

Second, the variograms presented in figures 7.45 to 7.48 show that the OLS residuals have a different correlation structure to the OLS residuals for G0 (presented in section 7.6.1). This result needs to be interpreted with some caution, since the number of pixels used to compute the variogram is substantially smaller for G0 – particularly when the variogram is calculated on a per-GT basis. This smaller sample size may lead to less stable variograms. Nevertheless, the variograms have a different form. In particular, the long range spatial structure evident for G0 for the visible wavebands is much less apparent for G1. Furthermore, the variogram for the concrete surface is steeper.

A consequence of the different spatial structure for G1, combined with the different weighting structure, was that the WML approach yielded a plausible numerical solution for all four wavebands with a variogram range of between 3 and 6 m. Combined with the different weighting structure, this led to much wider confidence intervals on the regression

line (see tables 7.28 to 7.28). These exceeded 2% (reflectance) for MMR bands 2 and 4 and were greater than 1% for bands 1 and 3. Indeed, for band 2 $\hat{y}$ differences in excess of 2.5% arose between OLS, WLS and WML. This is indicative of substantial uncertainty in the ELM for this waveband.

The results discussed above present a potentially confusing message. By only including pixels that contain two or more field measurements, it is intended that the accuracy of the per-pixel reflectance values should be *increased* (Curran & Williamson 1985). However, when these values are carried through into the ELM regression and the weighting and correlation structure modelled, the uncertainty actually *increases*. This situation was discussed in section 1.2 and arises when an increase in *understanding* leads to an increase in uncertainty. In this particular case "understanding" can be taken to mean more accurate input data and the use of a more appropriate model (i.e., one that includes weighting and correlation). An unfortunate consequence of this is that it is not possible to make clear recommendations for the practitioner in terms of specifying a sample size or sample structure that will reduce uncertainty. Whilst it is possible to decrease uncertainty in the regression by using G0 rather than G1 it is now known that this may not be the right thing to do. Hence, the high levels of uncertainty associated with G1 are something that the user must accept. Nevertheless, this does open two key areas for future research.

First, it might be argued that the total number of data points in G1 (111) is low for geostatistical analysis and the estimate of the correlation and weighting structure may be inaccurate. Future research could adopt a sampling strategy that would yield more pixels containing two or more field samples. This would allow investigation of the implications of sample size.

Second, if figure 4.4 is referred to, it is clear that the spectral response of MMR band 1 closely matches to ATM band 2. However, the matches between the other MMR bands and ATM bands is approximate. In particular, MMR band 4 is much wider than ATM band 7. It is possible that this may affect the regression model. This is most likely to be manifested through the grass surface, owing to the distinct spectral signature of vegetation by comparison to the relatively bland asphalt and concrete surfaces. In practice, addressing this problem is not straightforward. A spectroradiometer could be used in preference to the MMR and the spectra convolved with the ATM or CASI spectral response functions. However, such instruments are substantially more complex and more cumbersome than the MMR and it was not possible to obtain such a large sample. There are serious practical and technical difficulties to be addressed if a fine resolution spectral and spatial sample are to be obtained.

	MMR B1 on ATM B2	MMR B2 on ATM B3	MMR B3 on ATM B5	MMR B4 on ATM B7
$\hat{\beta}_0$	-2.5749	0.5676	-1.7292	-3.9062
$s.e.(\hat{\beta}_0)$	0.1306	0.3561	0.2159	0.3758
p	< 0.01	> 0.1	< 0.01	< 0.01
$p < 0.05$	Yes	No	Yes	Yes
$p < 0.01$	Yes	No	Yes	Yes
$\hat{\beta}_1$	0.3079	0.2547	0.35	0.5508
$s.e.(\hat{\beta}_1)$	0.0041	0.0104	0.0063	0.0097
p	< 0.01	< 0.01	< 0.01	< 0.01
$p < 0.05$	Yes	Yes	Yes	Yes
$p < 0.01$	Yes	Yes	Yes	Yes
$\hat{\sigma}^2$	1.0174	1.5791	0.9691	1.1158
$\hat{\sigma}$	1.0087	1.2566	0.9844	1.0563
Correlation model	Spherical	Spherical	Spherical	Spherical
$\hat{\phi}$	4.3304	6.2444	3.3253	6.1526
$\hat{\nu}$	0.3746	0.8742	1	0.925
R^2	0.9553	0.9417	0.9796	0.9856
$w_{asphalt}$	24.0837	21.5765	17.8257	6.6392
$w_{concrete}$	0.7152	0.5626	0.524	0.4645
w_{grass}	12.0368	8.5076	4.0467	0.5777
$w^{relative}_{asphalt}$	1.0000	1.0000	1.0000	1.0000
$w^{relative}_{concrete}$	0.0297	0.0261	0.0294	0.0700
$w^{relative}_{grass}$	0.4998	0.3943	0.2270	0.0870

TABLE 7.24: Table showing the result of parameter estimation of the ELM regression model using the simple averaging approach and the combined heteroskedastic and correlated residuals (WML) approach. Only pixels containing *two or more* (G1) field measurement were included in the regression model. The parameters were estimated estimated using the maximum likelihood approach with a variogram embedded to account for the spatial correlation.

Radiance	$\hat{y}_{OLS}$	CI	$\hat{y}_{WLS}$	CI	$\hat{y}_{WML}$	CI	Example Landcover
0	-2.34	0.69	-2.58	0.42	-2.57	0.52	-
10	0.66	0.54	0.50	0.30	0.50	0.36	Dense vegetation
20	3.66	0.40	3.58	0.18	3.58	0.22	Grass
30	6.65	0.30	6.66	0.11	6.66	0.14	Grass
40	9.65	0.27	9.74	0.16	9.74	0.20	Urban surfaces (e.g., asphalt)
50	12.65	0.34	12.82	0.27	12.82	0.33	Urban surfaces
60	15.65	0.46	15.90	0.39	15.90	0.49	Concrete
70	18.64	0.61	18.98	0.52	18.98	0.64	Concrete
80	21.64	0.76	22.06	0.65	22.06	0.80	Metallic surfaces (e.g., sheds)
90	24.64	0.92	25.14	0.78	25.14	0.96	Metallic surfaces (e.g., sheds)
100	27.63	1.08	28.22	0.91	28.22	1.12	Sand

TABLE 7.25: Table showing the predicted values of reflectance for MMR band 1 (regressed on ATM band 2) for a range of radiance values. Results are shown for the OLS, WLS and the WML type regression model. The reflectance data were aggregated using the simple averaging approach and only pixels containing *two or more* (G1) field measurement were included. The 95% confidence interval about the predicted values are also shown.

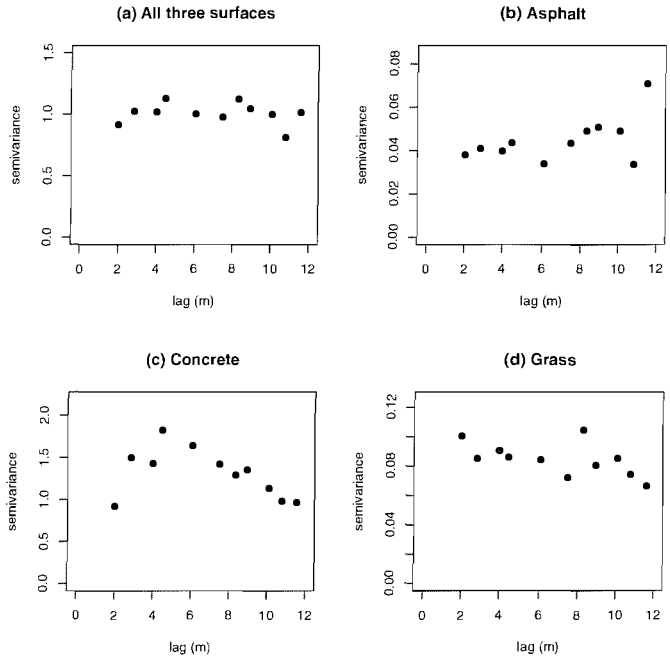


FIGURE 7.45: Plots showing the residuals from the OLS regression model for MMR band 1 regressed on ATM band 2. The MMR data were aggregated using the simple averaging approach, where only pixels containing *two or more* (G1) field measurements were retained. The variogram for all three surfaces (a) were calculated for the standardardised residuals. The variograms for the individual GTs were calculated using the non-standardised residuals.

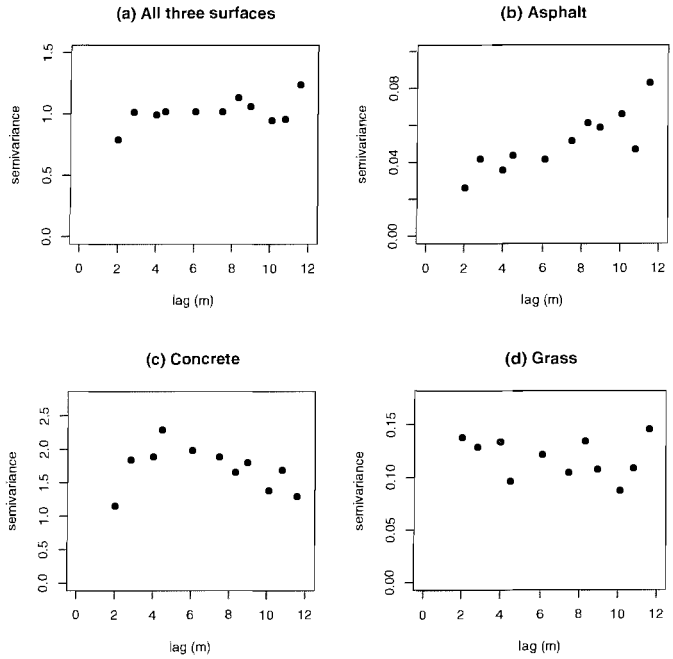


FIGURE 7.46: Plots showing the residuals from the OLS regression model for MMR band 2 regressed on ATM band 3. Other details are as for figure 7.45

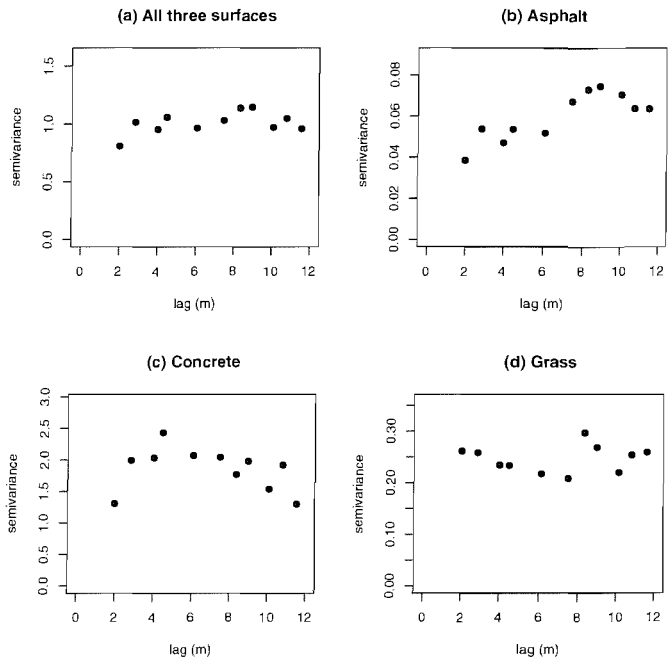


FIGURE 7.47: Plots showing the residuals from the OLS regression model for MMR band 3 regressed on ATM band 5. Other details are as for figure 7.45

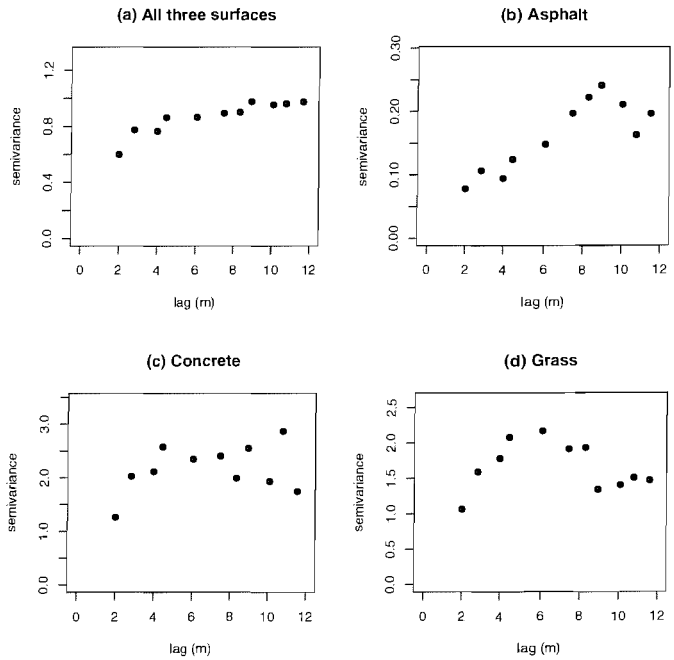


FIGURE 7.48: Plots showing the residuals from the OLS regression model for MMR band 4 regressed on ATM band 7. Other details are as for figure 7.45

Radiance	$\hat{y}_{OLS}$	CI	$\hat{y}_{WLS}$	CI	$\hat{y}_{WML}$	CI	Example Landcover
0	-0.42	0.82	0.36	0.83	0.57	1.41	-
10	2.48	0.67	2.97	0.60	3.11	1.01	Dense vegetation
20	5.39	0.52	5.57	0.37	5.66	0.63	Grass
30	8.29	0.40	8.17	0.19	8.21	0.31	Grass
40	11.19	0.33	10.77	0.21	10.75	0.38	Urban surfaces (e.g., asphalt)
50	14.10	0.34	13.37	0.41	13.30	0.72	Urban surfaces
60	17.00	0.43	15.97	0.63	15.85	1.11	Urban surfaces
70	19.90	0.55	18.57	0.87	18.39	1.51	Concrete
80	22.81	0.70	21.17	1.10	20.94	1.92	Concrete
90	25.71	0.85	23.77	1.34	23.49	2.33	Sand
100	28.61	1.01	26.38	1.57	26.03	2.73	Metallic surfaces (e.g., sheds)

TABLE 7.26: Table showing the predicted values of reflectance for MMR band 2 (regressed on ATM band 3) for a range of radiance values. Details are given in table 7.25.

Radiance	$\hat{y}_{OLS}$	CI	$\hat{y}_{WLS}$	CI	$\hat{y}_{WML}$	CI	Example Landcover
0	-1.69	0.76	-1.72	0.67	-1.73	0.86	-
10	1.81	0.61	1.78	0.48	1.77	0.61	Dense vegetation
20	5.32	0.48	5.28	0.30	5.27	0.38	Grass
30	8.82	0.38	8.77	0.15	8.77	0.20	Urban surfaces (e.g., asphalt)
40	12.33	0.32	12.27	0.19	12.27	0.25	Urban surfaces
50	15.83	0.34	15.77	0.35	15.77	0.46	Urban surfaces
60	19.34	0.43	19.26	0.54	19.27	0.70	Concrete
70	22.84	0.55	22.76	0.73	22.77	0.94	Concrete
80	26.35	0.69	26.26	0.92	26.27	1.19	Sand
90	29.85	0.84	29.75	1.12	29.77	1.44	Metallic surfaces (e.g., sheds)
100	33.36	0.99	33.25	1.31	33.28	1.69	Metallic surfaces (e.g., sheds)

TABLE 7.27: Table showing the predicted values of reflectance for MMR band 3 (regressed on ATM band 5) for a range of radiance values. Details are given in table 7.25.

Radiance	$\hat{y}_{OLS}$	CI	$\hat{y}_{WLS}$	CI	$\hat{y}_{WML}$	CI	Example Landcover
0	-3.86	1.68	-3.82	0.82	-3.91	1.49	-
10	1.60	1.40	1.64	0.62	1.60	1.13	Water
20	7.05	1.12	7.10	0.43	7.11	0.80	Shallow water
30	12.50	0.87	12.56	0.29	12.62	0.54	Urban surfaces (e.g., asphalt)
40	17.96	0.64	18.02	0.26	18.12	0.49	Urban surfaces
50	23.41	0.50	23.48	0.38	23.63	0.69	Concrete
60	28.87	0.51	28.94	0.56	29.14	1.01	Concrete
70	34.32	0.67	34.40	0.76	34.65	1.36	Grass
80	39.77	0.90	39.86	0.96	40.16	1.72	Grass
90	45.23	1.17	45.32	1.17	45.66	2.10	Dense vegetation
100	50.68	1.44	50.78	1.37	51.17	2.47	Dense vegetation

TABLE 7.28: Table showing the predicted values of reflectance for MMR band 4 (regressed on ATM band 7) for a range of radiance values. Details are given in table 7.25.

7.7 Summary and advice for the practitioner

This chapter began by discussing the problem of heteroskedasticity and introduced the WLS approach to tackling it. The WLS approach is straightforward and introduced in most statistics textbooks (Sen & Srivastava 1990, Neter et al. 1996, Gelman et al. 2004). However, the difficulty in applying WLS lies in accurately specifying the weights. This seems to be universally understated in textbooks. The approach taken in section 7.2 was to divide the OLS residuals into batches according to their GT. The per-GT weight was then given as the inverse of the variance of the OLS residuals. This follows the approach suggested by Neter et al. (1996) and Gelman et al. (2004). This approach was effective for all four MMR wavebands.

The primary drawback of the above approach comes when predicting reflectance over a surface that is not included in the training data. This is because the above approach does not provide a weight for such surfaces. Hence four alternative strategies for *predicting* the weight were examined. Of the four strategies investigated, strategy 3 was most effective for the visible wavebands. However, it is necessary to recognise the caveat that this does not guarantee success for unmeasured surfaces. An effective strategy was not found for the NIR waveband (MMR band 4).

Previous discussion has advanced the argument that the ELM could be considered a form of calibration. Hence, primary interest should lie in the regression coefficients and the confidence intervals surrounding the regression line rather than in the specification of prediction intervals. For the purpose of the ELM, it might be argued that weights based on the inverse of the variance of the OLS residuals are sufficient. However, part of the aim of this chapter was to examine methodologies that would be more generally applicable. In other GIS and remote sensing applications of regression, it would be necessary to calculate prediction intervals – hence predicting the w_i over an unmeasured surface would be necessary. However, the analysis presented in section 7.3 demonstrated the difficulty of specifying a model to *predict* the weights. In particular, an appropriate model was not found for the NIR waveband and the strategy (3) adopted for the visible wavebands was less accurate than simply dividing the OLS residuals into batches. The difficulty attached to specifying and predicting weights appears to be universally ignored in textbooks and in remote sensing and GIS journal papers. However, this analysis does emphasise that, where prediction is the goal, the user should establish a mechanism for empirically calculating the weights over surfaces of interest.

The second part of the chapter examined a geostatistical method for conducting regression where the residuals are spatially correlated. Mathematically this is identical to the ML and REML approaches for model-based geostatistics (MBG). However, it was argued in sections 2.2.2.4 and 2.4 that the emphasis of geostatistics is interpolation rather than prediction on the basis of covariates (Dungan 1998, Lark 2000). Hence, applications of geostatistics tend to give limited attention to the estimation of the regression coefficients.

Nevertheless, some examination of the ML approach is given in the literature in the context of regression (Harris & Johnson 1996, Lark 2000, Ver Hoef et al. 2001, Lark & Cullis 2004). These authors emphasise the importance of modelling the correlation function, both for accurate parameter estimation and for hypothesis testing. In particular, ignoring the correlation structure can lead to incorrectly incorporating covariates into regression models (Harris & Johnson 1996, Ver Hoef et al. 2001, Lark & Cullis 2004).

Lark (2000) is unique in the use of a simulation study to investigate the effectiveness of the ML geostatistical regression model. However, Lark focused primarily on the estimation of σ^2 rather than on parameter estimation more generally. Furthermore, the simulated data were set up with very strong spatial structure and his model did not include a nugget component. The simulation used in this thesis had a weaker spatial structure and the model incorporated a nugget component. The ML and REML models were investigated. REML was found to give slightly less biased estimates of σ^2 and was larger than the OLS estimate, as expected from theory. Both approaches gave unbiased estimates of ϕ. Estimates of ν were both biased and imprecise. The effect of the correlation structure on the confidence intervals on β was also shown. It was demonstrated that confidence in the estimation of β_0 decreased, whilst confidence in the estimation of β_1 increased. Hence, there is not a straightforward relationship between the value of σ^2 and the width of a confidence interval. This is not explored by Lark, although some limited attention is given by Harris & Johnson (1996).

When the sample size was increased (3×100) both ML and REML gave accurate estimates of σ^2, ν and ϕ for the structured sample. However, it is important and interesting to note that, for the randomly located samples, the model was able to yield approximate estimates of ϕ and ν. This shows that the model is still "detecting" evidence of spatial structure in a random sample. Hence, simply by virtue of sampling randomly, independence *cannot* be assumed.

Finally, the model was implemented with variable values of σ^2 within a combined weighted-ML model. For known weights, this was found to be effective for simulated data.

Overall, the research conducted using the simulated data showed that the ML and weighted-ML model could be effectively applied to regression with spatially correlated residuals. However, the true test of the approach has to lie with its application to real data. This was undertaken for the ELM and the "G0" sample in section 7.6.1. The variograms of the OLS residuals showed clear evidence of spatial structure in all four wavebands. However, the pattern of this spatial structure was not straightforward – particularly for the visible wavebands. Furthermore, the nature of the spatial structure showed evidence of varying between GTs. This raises the question of whether a single value of ν and ϕ can be applied across all GTs. The combined weighted-ML model was not successful for the visible wavebands, although it was for the NIR waveband (MMR band 4). Importantly, incorporation of the correlation structure led to an *increase* in the width of the confidence interval on the regression line relative to the OLS and WLS cases. This is an important result, because

it shows that failure to account for heteroskedasticity leads to an *underestimation* in the uncertainty in the ELM regression line.

As noted above, the WML approach failed for the visible wavebands. Hence, an *ad hoc* implementation of the RK approach was adopted where the correlation structure was modelled from the OLS residuals for each waveband. This was used to estimate the GLS parameters for both the visible and NIR wavebands. For the visible wavebands, the confidence intervals were wider than for the OLS or WLS cases. In particular for MMR band 2, this led to a change in predicted values of y of approximately 1% reflectance. These results emphasise the importance of accounting for the correlation and weighting structure. However, the RK approach was not examined in detail in this thesis. A further examination of the RK approach, using simulated data, is a potential avenue for future research.

Finally, section 7.6.2 discussed the implications of using the G1 sample rather than the G0 sample. When the G0 sample was used the uncertainty in the parameter estimates and the regression line increased. This was particularly apparent for MMR bands 2 and 4 and led to a change of up to 2% reflectance in $\hat{y}$ as well as a widening of the confidence intervals. This is attributed to a change in $Cov(\hat{\beta})$, a change in the weighting structure and a change in the correlation structure between G0 and G1. Future research will consider the sample size as well as the spectral response of the MMR and ATM instruments.

In summary, the results presented in this chapter have demonstrated the importance of accounting for heteroskedasticity and spatial correlation when implementing the regression model. For the ELM, failure to account for these can lead to inaccurate estimations of the regression coefficients as well as uncertainty in the location of the regression line, as manifested by the confidence intervals. In the more general setting both heteroskedasticity and spatial correlation would be highly important for prediction and for quantifying uncertainty in prediction. However, the analysis conducted for this chapter highlighted the difficulty of using "real world" data to perform complex statistical inference (Diggle et al. 1998). This was clearly demonstrated for the NIR waveband (MMR band 4), for which no satisfactory approach was found for *predicting* the weights for unsampled surfaces. It was also demonstrated for the visible wavebands, where the weighted-ML approach failed for the G0 sample.

Chapter 8

Discussion

The key issues that were investigated in this thesis were stated in the aims and objectives given in section 1.3. These were the subject of the analysis in chapters 5, 6 and 7. The objective of this chapter to summarise and discuss the major findings. Where appropriate a more general discussion of the implications for empirical modelling in remote sensing and GIS is provided. The discussion is grouped under three headings:

1. *Data pairing.* This section discusses the mechanisms for combining remotely sensed and field data, as discussed in chapter 5.
2. *Positional uncertainty.* This section develops the discussion of positional uncertainty, which was presented in chapter 6.
3. *Spatial modelling.* The issues of spatial modelling, raised in chapter 7, are considered.

8.1 Data pairing

Chapter 5 began by examining the typical approach to implementation of the ELM. This approach was shown to be problematic for various reasons:

1. The number of data pairs used for regression modelling is reduced to the number of targets. In chapter 5 this meant that three data pairs were used (four for the modified-typical approach). This led to a low number of degrees of freedom for estimating the regression parameters and low confidence in the estimates of the regression parameters. Furthermore, the approach was sensitive to sample size.
2. The typical approach made inefficient use of the field data. Substantial effort was expended to spatially sample each GT. However, much of this information was then "thrown away".

3. The field-based reflectance measurements showed a bimodal distribution for the concrete target. Spatial summary plots showed a clear patch of relatively bright measurements. This is indicative of a target of mixed composition.

4. The reflectance and radiance measurements were defined on ambiguous spatial supports of approximately 30 m × 30 m. However, prediction was required on pixel-sized supports. The base pixel size for this imagery is 2 m. It is well known that relationships defined on one support may not be valid at another support (see section 2.2.1). Furthermore, even if the relationship holds the variance tends to decrease as the variables defined on point supports are aggregated to larger ones.

These limitations were addressed by defining the data pairs on a per-pixel rather than per-target basis. This addresses the first two issues by making full use of the data set and by increasing the number of data pairs used to estimate the parameters of the regression model. The third issue is partly addressed since within-pixel variability rather than within-target variability is now important. Defining the data pairs on a per-pixel, rather than per-GT basis gives the user greater flexibility in his or her choice of GT. The user should still take care when choosing ground targets. It would clearly be unwise to choose a small bright target adjacent to a much darker surface owing to the adjacency effects described in sections 3.1.1 and 3.2.3.2. The contrast with the concrete GT is clearly far less extreme, although the potential impacts of this were not examined.

Defining the data pairs on pixel-sized supports also allowed the fourth issue to be addressed since the model is parameterised on the same support as that required for prediction. This relates to the modifiable areal unit problem (MAUP), where the relationship between variables may change according to the level of *aggregation* and to the *zonation* employed (Fotheringham & Wong 1991, Gotway & Young 2002) (see also section 2.2.1). In principle, zonation effects could be examined by aggregating the 2 m pixels in different ways. For example, if a 2 × 2 aggregation is performed, there are four different starting points for aggregating from 2 m to 4 m square pixels. This effect was not examined in this thesis, although it is expected that it would be less important for GTs than for highly heterogeneous surfaces. Hence only aggregation, or change-of-support, effects were considered.

For the full sample size, the point-pixel and block-pixel approaches presented in chapter 5 all yielded similar estimates for the regression coefficients, although the uncertainty about the estimates and predictions did vary. The estimated coefficients for the typical and modified-typical approaches indicated a slightly steeper regression line. This slight steepening of the regression line was also observed for the point-pixel approach, when the 2 m pixels were aggregated to 4 m and 6 m pixels in section 6.2.3 (see figures 6.5(a) and 6.6(a)), although this phenomenon was not observed for 8 m pixels. When the sample size was reduced the sample-to-sample variability was greater for the typical approach than for the point-pixel approach. For the typical approach, the effect of the change-of-support was conflated with sample size and the bimodality of the concrete surface. When

the sample size was reduced this led to greatly increased uncertainty in the estimate of the mean of the concrete target. When inference was undertaken at pixel-sized supports the problem of within GT variability was eliminated and replaced with the necessity to have an unbiased estimate of reflectance on a per-pixel basis. However, it was still necessary to sample a sufficiently large number of pixels to minimise sample-to-sample variability.

Chapter 7 modelled the heteroskedasticity and spatial correlation among the residuals. The incorporation of these affects *both* the regression parameters and the uncertainty in estimation and prediction. It is well established that increasing the size of the support of a variable leads to a decrease in the variance (see section 2.2.1). If this effect differs between GTs then it will affect the relative weights used in the regression model. Furthermore, increasing the size of support leads to a decrease in the variogram (Atkinson & Curran 1995, Webster & Oliver 2001). Hence the spatial structure in the residuals will vary between different pixel sizes. As a result the pixel size may affect both parameter and uncertainty estimates for WLS and WML approaches. This effect was not examined in this thesis so the extent of any changes cannot be evaluated.

There were two broad alternatives offered to the typical approach. The first was the point-pixel approach and the second was the block-pixel approach. The first did not fully address the support issue, since the predictor variable is defined on a pixel support and the reflectance measurements are defined over small areas. However, the support mismatch was lower than for the typical approach. Furthermore, this approach was straightforward to implement, given information on the position of each field measurement. It also meant that all the information content in the data was retained. Utilisation of this approach yielded regression models with tight confidence intervals. It was recommended that *at least* 50 samples be collected per target, with a minimum of 100 being the preferred number.

The next step was to implement the block-pixel approach, whereby the field measurements were aggregated to pixel-sized blocks and then paired with the co-located pixels. Two approaches for aggregation were utilised – geostatistical block kriging and simple averaging. Kriging provides an optimal interpolation of a spatial sample and is a widely implemented mechanism for aggregation (Heuvelink & Pebesma 1999, Bierkens et al. 2000, Gotway & Young 2002, Banerjee et al. 2004). However, it was problematic for the implementation of the ELM for two reasons. First, although the variography showed clear evidence for spatial structure, the kriged predictions were inaccurate. The accuracy could be increased by taking a denser sample although logistic constraints mitigate against this. Second, there are theoretical problems associated with combining kriged or conditionally simulated surfaces in regression (Atkinson & Kelly 1997, Atkinson & Tate 2000), as discussed in sections 2.2.2.6 and 5.4.3. Finally, the geostatistical block kriging approach is complex. Many users of remotely sensed data are unlikely to want to implement it (Atkinson 2005).

The alternative to block kriging was the simple averaging approach to aggregation. This is of comparable simplicity to the point-pixel approach, except all field measurements

contained within a given pixel are averaged rather than being treated individually. Under this framework it is assumed that the average for each pixel is an unbiased estimate of the true value, regardless of the number of points that it is averaged over. Relative to the point-pixel approach, this method provides a more accurate estimate of the true reflectance for each pixel (Curran & Williamson 1985). If pixels containing two or more reflectance measurements were retained, this gave a total of 111 data pairs (36 or 39 for each GT) for the ELM regression. This delivered regression models with tight confidence intervals for OLS. However, this sample size was considered small for parameter estimation of subsequent spatial regression models (chapter 7). Hence, the additional pixels that contained only one field measurement were also utilised for analysis.

It is recommended that the simple averaging approach be used to characterise pixels within GTs for operational use of the ELM. However, the sample size and strategy requires further consideration. This should be designed in such a way that sufficient measurements are obtained to accurately characterise each pixel (bearing in mind that the pixel boundaries are not known *a priori*) while also being of sufficient spatial extent to cover several pixels.

The importance of the number of field measurements required to adequately characterise a pixel became clear when there was heteroskedasticity and correlation amongst the regression residuals. This was demonstrated in section 7.6.2 when the results of implementing the simple averaging approach with one-or-more or two-or-more field measurements per pixel were compared. Averaging over different numbers of measurements per-pixel had consequences for calculation of the relative weights and altered the form of the spatial structure amongst the residuals. This impacted both on the estimates of the regression coefficients and on the uncertainty, as quantified by the confidence intervals around the regression line. Both of these, ultimately, affected prediction. The implication of the results presented in section 7.6.2 is that *at least* two reflectance measurements are required for each pixel. Further research is required to test whether obtaining more than two measurements per-pixel leads to substantial changes in estimation and prediction.

8.2 Positional uncertainty

The alternatives to the typical implementation of the ELM, summarised in the previous section, were contingent on knowing the location of the field measurements – both relative to each other and within the image.

Where remotely sensed and field data are combined in empirical models it seems self evident that the variables should be co-located, since positional uncertainty may lead to field data being combined with the "wrong" pixel (Gao 2006). However, although this issue has received some comment in the literature (Curran & Hay 1986, Prince & Astle 1986, Larsson 1993, Dungan 2002, Oudemans et al. 2002, Zha et al. 2003, Gao 2006), these papers do not quantify the impact or present solutions to the problem. Some authors have sought to limit the effects of positional uncertainty by averaging over several

pixels (e.g., Justice & Townshend 1981, Li et al. 1998, Smith & Milton 1999). However, this approach does not quantify the issue or provide a model based solution. This also raises concern that the support of the two sets of measurements will not be matched, as discussed by Dungan (2002) and in the previous section.

For the block-pixel approach, field data were aggregated to the pixel-sized support prior to combining them with the remotely sensed data. In this case positional uncertainty in the field measurements may affect the accuracy with which the per-pixel value of reflectance is estimated. More generally, positional uncertainty affects the accuracy of geostatistical estimation and prediction, as discussed in sections 2.3 and 6.3.

For the block-pixel approach, introducing substantial positional uncertainty ($\zeta = N(0, 4)$, led to small, but clear, bias in the estimate of the regression coefficients. However, the estimate of the residual variance, σ^2, increased from 0.77 to approximately 1.2 for the full data set. This led to a widening of the confidence and prediction intervals around the regression line. The bias in the regression coefficients was small when greater than 50 measurements were taken per target although the bias in the estimate of σ^2 remained. It was recommended that *at least* 50 measurements be taken for the 2 m pixels and 30 measurements for pixels that are 4 m or larger. However, to limit the effect of bias in σ^2, 100 measurements per target are preferable for all pixel sizes. These are the same sample sizes that were recommended in the absence of positional uncertainty in order to reduce sample-to-sample variability (see section 5.3.1).

For the geostatistical block-pixel approach the introduction of positional uncertainty led to an increase in the variogram at short lags, as expected from theory (Atkinson 1996) and from simulations (Atkinson 1996, Gabrosek & Cressie 2002). The exception was the grass surface, where the variogram was reduced at all lags. The reason for this is unclear, although it was proposed that this may be due to inaccurate modelling of the spatial structure. Importantly, introducing positional uncertainty did not, ultimately, affect substantially parameter estimation for the ELM. Whilst this is an encouraging result, it needs to be interpreted in the context of other problems associated with using block kriged estimates in regression models. Further discussion is given in section 6.4.

Overall the results of the analysis should be encouraging for users of the ELM since, even when the positional uncertainty is large relative to the pixel size, the impact on parameter estimation is low providing a large sample is used. However, this should *not* be taken as reason to take little care about recording location information, since the estimate of the residual variance still increases for all sample sizes for the block-pixel approach.

Section 6.4 gave several general suggestions for future research on positional uncertainty. Two of these are highlighted here. First, for the point-pixel approach, the analysis isolated positional uncertainty from sample variability. It is not clear how these two factors will interact and this remains a subject for future investigation. Second, the effect of positional uncertainty on the simple averaging approach was not examined. This approach, by definition, yields an average value of reflectance for each pixel. Hence, it is expected

that the effect of positional uncertainty will not be worse than it was for the point-pixel approach. This assertion requires further empirical evaluation.

The extent to which these results are generalisable to other applications of empirical models is unclear. In particular, the impact of positional uncertainty depends on the strength of correlation between the variables, the heterogeneity of the surface and the magnitude of the attribute uncertainty attributable to positional uncertainty relative to the overall variability of the same variable (Salvador 1999). When utilising the ELM, explicit attention is given to selecting GTs that are relatively homogeneous. This may not be the case for other applications of empirical models. Hence the impact of positional uncertainty may, potentially, be larger (e.g., Salvador 1999).

8.3 Spatial modelling

Examination of the diagnostic plots for the point-pixel and block-pixel approaches showed that the residuals were clearly heteroskedastic. This situation was addressed by using weighted least squares (WLS) regression (Sen & Srivastava 1990, Neter et al. 1996), where the weights were given as the inverse of the residual variance for each GT. This approach was satisfactory for addressing the problem of heteroskedasticity. The WLS approach was satisfactory for parameter estimation and for calculating accurate confidence intervals. However, if the prediction uncertainty is required it is unsatisfactory. This is because it provides weights *only* for the GTs that are used in the parameter estimation stage but not at locations that are not used for estimation. Alternative approaches were examined that could be used to predict the weights given the local variance (LV) for a 3×3 window surrounding the pixel of interest. To do this, the OLS residual was regressed on LV and the possibility of specifying weights on a per-pixel, per-GT and per-(landcover) class were examined. The diagnostic plots showed that the per-pixel and per-GT approaches were effective at reducing heteroskedasticity for the visible MMR wavebands, but not for the NIR MMR waveband. The per-class approach was not successful due to within-class variability. Overall, this analysis emphasised the difficulty in predicting appropriate weights for unsampled targets. This was possible for the visible wavebands (although this is no guarantee that these will be accurate for unsampled targets), but not for the NIR waveband. The user should proceed with caution when predicting weights for unsampled targets. It would be preferable to obtain empirical evidence for the weights over particular targets of interest.

In addition to the problem of heteroskedasticity, variograms of the OLS residuals showed them to be spatially correlated. The geostatistical approach to handling spatially correlated residuals in regression has received some attention in the literature (Cook & Pocock 1983, Harris & Johnson 1996, Lark 2000, Ver Hoef et al. 2001, Lark & Cullis 2004). However, the presentation is largely theoretical and tests on simulated data have been limited (e.g., Lark 2000). Furthermore, it has not been extended to deal with heteroskedastic

residuals. Hence, the approach was tested first with simulations and was shown to provide accurate estimates of the variance parameter (σ^2) and variogram range (ϕ) but less accurate estimates of the proportion of spatially structured variance (ν) for a 100×100 grid. However, the accuracy of the estimation of all three parameters increased for a large sample of $3 \times 100 \times 100$ grids. An interesting and important result was observed for this large sample since the ML approach "detected" spatial structure even when a random sample was used. The implication of this is that choosing a random sampling design will not, necessarily, lead to a model with independent errors and the user should address this. Finally, the approach was shown to work for regression models with heteroskedastic residuals when appropriate weights were provided.

The weighted-ML (WML) model for spatially correlated heteroskedastic residuals was then applied to the ELM data, where the pixel-based radiance data were paired with simple-averaged blocks. The three regression models: OLS (ordinary least squares), WLS (weighted least squares) and the WML (weighted maximum likelihood), were then compared. WML models explicitly both the spatial correlation and the heteroskedasticity that are known to occur. By applying this model the structural uncertainty (see section 1.2) is reduced relative to OLS and WLS. Hence, in the absence of further mitigating factors, the parameter estimates and associated confidence intervals are most appropriate.

The simple-averaged blocks were constructed either by retaining all blocks that contained one or more field measurements (the G0 sample) or by omitting all blocks containing fewer than two field measurements (the G1 sample). Although the G1 sample gives more accurate per-block estimates of reflectance, the analysis was performed first using the G0 sample owing to its relatively large size (436 *vs.* 111).

For the NIR waveband (MMR band 4 regressed on ATM band 7), the WML approach yielded reflectance predictions that were up to 0.7% different, by comparison to OLS. By comparison to WLS, the WML model yielded wider confidence intervals around the regression line. By comparison to OLS, the confidence intervals were wider for brighter pixels but narrower for darker ones. For vegetated surfaces, which often interest practitioners, the 95% confidence intervals varied between 0.7 and 1.3% reflectance.

When the WML model was applied to the visible wavebands, implementation of the model was unsatisfactory. A limitation of the WML approach was that, although it accounts for variability in σ^2 between GTs, the nugget and range are fixed for all GTs. Assessment of the residual variograms for each target showed that this was not the case. Nevertheless, the objective was to provide a general approach that could also be used for estimation and prediction across a range of GTs and landcover types, although it might be suboptimal for given cover type. Ultimately, this was not tenable and, although the WML model was applied successfully for simulated data and for the NIR waveband, it was not appropriate for the visible wavebands where the surfaces exhibited a more complex spatial structure. It is unclear whether an elegant solution can be found to this problem. However, an ad hoc regression kriging (RK) approach was applied, where the correlation matrix was

estimated using the OLS residual variogram for each GT. For MMR bands 1 and 3 the predicted values of y were comparable to those obtained using OLS and WLS, although substantial differences were found for MMR band 2. The 95% confidence intervals were up to 0.5% reflectance wider than those obtained from OLS. This RK approach did allow incorporation of the correlation structure. However, it is ad hoc by comparison to the WML approach and this requires more detailed consideration. Furthermore, although this approach works for estimating the regression parameters, it could not be used for prediction since it would give no information about the spatial structure of targets that that were not used for parameter estimation.

Finally, the results obtained for the G0 sample were compared to those obtained for the G1 sample. This led to comparable parameter estimates and prediction of $\hat{y}$ for OLS, although the confidence intervals were wider. However, the situation was more complex for the WLS and WML approaches. In particular, for WML, this led to a steepening of the regression line for MMR bands 1 and 4. This led to increases in $\hat{y}$ of up to 0.7% reflectance for band 1 and 1.6% reflectance relative to the G0 case. Predictions obtained for MMR band 2 were comparable to those obtained using the G0 sample. For WML the 95% confidence intervals were wider than the G0 case for these three wavebands, being up to 0.2% wider for band 1, 0.7% for band 3 and 1.2% for band 4. The greatest difference was observed for MMR band 2, which showed differences in excess of 2% in $\hat{y}$ and confidence intervals up to 1.8% wider.

Discussion of the relative merits of using the G0 and G1 sample was given in section 7.6.2. In particular, whilst it is possible to decrease the width of the confidence intervals by using G0 rather than G1, it is now known that this may not be the correct thing to do. The user must accept this uncertainty. Furthermore, the width of the confidence interval varies according to reflectance of the surface, which is linked to landcover type, and waveband. Hence the impact of uncertainty will depend on the application. If the user finds that this is too great then they can either change the data or change the model. However, neither of these is guaranteed to lead to greater certainty in the model output.

Chapter 9

Conclusions

This chapter highlights the key findings and recommendations of the research and emphasises their significance for application of the ELM and for empirical modelling in remote sensing more generally. Avenues for future research are then discussed. Finally, a concluding statement is offered.

Section 1.3 outlined the aims and objectives for the research. It was stated that, although conceptually straightforward, problems with the ELM arise owing to uncertainties in (i) standard practice (the typical approach), (ii) the data and (iii) the model. Typical implementation of the ELM led to wide confidence intervals on the regression line, even when the sample size was very high (approximately 200 measurements per GT). This was exacerbated when the sample size was reduced. The typical approach to the ELM is very widely used. It is important for users to be aware of these limitations and to understand that the uncertainty in this implementation of the ELM is very large.

The problems with the ELM led to the formulation of three key issues that were examined in this thesis:

- Pairing the remotely sensed and field data.
- The effect of positional uncertainty in field data on the ELM.
- Modelling heteroskedasticity and spatial autocorrelation in the residuals.

9.1 Data pairing

Alternatives to the ELM were investigated whereby the field-based measurements of reflectance were (i) directly paired with their co-located pixels (the point-pixel approach) and (ii) aggregated to pixel-sized blocks (the block-pixel approach). Both these approaches made more effective use of the data and were improvements on the typical approach since they reduced uncertainty in the estimation of the regression line.

The point-pixel approach should be implemented with a minimum 50 field measurements per target, although at least 100 were recommended. This number is towards the high end of what is reported in the literature (Farrand et al. (1994): 100; Smith & Milton (1999): 5–35; Karpouzli & Malthus (2003): 15–60). The user should expect to obtain at least this number of field samples and they should record the location of these field measurements and document the number of measurements taken. This is not a trivial task, but is manageable given the availability of modern surveying and GPS equipment.

The block-pixel approach sought to reduce uncertainty over support by explicitly matching the support of the remotely sensed and field data and was preferred over the point-pixel approach for this reason. The simple averaging approach is recommended over the block kriging approach (see sections 2.2.2.6 and 5.4.3 for details). Implementation requires no more information than is required for the typical approach. Where pixels containing two or more field measurements are retained for the ELM the total number of measurements (111 blocks formed from 371 field measurements) is comparable to that recommended for the point-pixel approach.

9.2 Positional uncertainty

The point-pixel and block-pixel approaches were contingent on knowing the location of the field measurements. For the point-pixel approach, the effect of substantial positional uncertainty (standard deviation of 2 m, equal to the pixel size) was minimal for estimating the regression coefficients, although the estimated residual variance increased. The effect of positional uncertainty on the simple averaging approach should not be worse than for the point-pixel approach.

The level of positional uncertainty introduced is comparable to that obtained when using an inexpensive hand-held GPS operating in differential mode (Letham 2003). Given careful fieldwork and the use of scientific surveying or GPS equipment more accurate results can be expected. However, although the impact of positional uncertainty was limited, it is recommended that users should still give careful attention to the recording the location of field measurements. There are several unanswered questions relating to (i) the impact of positional uncertainty on the simple averaging approach and (ii) the extent to which positional uncertainty interacts with other factors such as sample-to-sample variability as well as the spatial modelling undertaken in chapter 7.

9.3 Modelling issues

Implementation of the block-pixel approach led to heteroskedasticity and spatial autocorrelation in the residuals from the regression model. Note that spatial autocorrelation is

distinguished from serial autocorrelation. Heteroskedasticity is modelled by applying appropriate weights. Specifying the weights is straightforward for parameter estimation and calculation of confidence intervals as required by the ELM. In more general applications of empirical models, it would also be necessary to calculate the weights at unsampled locations. However, it was not possible to develop a general model to predict the weights at unsampled locations. If these are required the user must determine them empirically.

A geostatistical model was combined with the weighting structure to model the combined effects of heteroskedasticity and correlation (the WML model). It was applied first for simulated data and then for the ELM. This extended previous research (Cook & Pocock 1983, Harris & Johnson 1996, Lark 2000) that focused on $\hat{\sigma}^2$ by (i) introducing the correlation structure and (ii) examining the accuracy of all parameter estimates. It was clearly demonstrated that spatial structure has wider implications than the estimation of σ^2. In particular it also affects $Cov(\hat{\beta})$ that in turn affects hypothesis testing and the calculation of confidence and prediction intervals – although the precise nature of this effect is not clear *a priori*. Furthermore, spatial structure was shown to affect parameter estimation even when the samples were gathered at random locations. Hence the effect of spatial structure cannot be eliminated by applying a specific sampling strategy. This effect would not be shown if only $\hat{\sigma}^2$ was examined.

Finally, the WML model was applied to the ELM where the reflectance measurements were aggregated using the simple averaging approach. These results were presented in sections 7.6.1 and 7.6.2 and are preferred over those presented in section 5.4.1. This is because, by modelling heteroskedasticity and spatial correlation, *structural uncertainty* in the model has been reduced. Hence the associated parameter estimates and confidence intervals are most legitimate. Confidence intervals indicate the likely range of estimated values, so wide parameter estimates are indicative of large uncertainty. This does not mean that the model with narrower confidence intervals should automatically be chosen since, as in this case, the more appropriate model may yield wider confidence intervals.

There was some debate as to whether simple–averaged blocks that included one-or-more (G0) or two-or-more (G1) field measurements should be included. The G0 sample is larger, which is desirable for geostatistical analysis. However, the G1 sample gives more accurate per-block estimates of reflectance and, for that reason, is considered most appropriate. The extent to which a larger G1 sample may lead to changes in the parameter estimates and confidence intervals remains a subject for future research.

9.4 Future research

This thesis advocates using the simple averaging approach for aggregating from field measurements to pixel sized blocks. However, the sampling scheme employed was inefficient since it led to several blocks that contained only one pixel. Future research should identify and test a sampling strategy that makes more effective use of the field measurements to

accurately estimate per-block reflectance. This could also be used to examine whether, given a larger sample size, uncertainty in the results presented section 7.6.2 can be reduced.

The problem of combining data that are defined on different spatial supports is widespread in remote sensing (Dungan 2002). The simple averaging approach may not be appropriate in situations when the support mismatch between field or remotely sensed data is very large or where it is not possible to obtain large numbers of field measurements. A more general approach would be to combine change-of-support and regression into one model (Gotway & Young 2002).

Surfaces have characteristic spectra, such that the reflectance in one band is not independent of reflectance in neighbouring bands. Indeed, the objective of atmospheric correction is to "correct" the whole spectrum as a whole (Clark et al. 2000). The research in this thesis modelled reflectance on a band-by-band basis and ignored the information in other bands. In principle, this information could be used to constrain the ELM regression. However, this is not straightforward because spectral profiles are surface dependent.

The research undertaken in this thesis focused on the spatial domain. However, reflectance of targets can vary over temporal scales ranging from hours to years due to wetting and drying, biophysical activity, changes in atmospheric conditions and BRDF effects (Moran et al. 2001, Anderson 2005, Anderson & Milton 2006). Future research could investigate how such changes are manifested in space. Furthermore, if temporal variability can be modelled it widens the time window over which field measurements can be taken (Moran et al. 2001) offering the opportunity to gather a larger data set.

Combined radiative transfer and ELM approaches have been advocated in the literature (Clark et al. 2000, Moran et al. 2001, Moran et al. 2003). Such models present an interesting topic for future research because it is not clear what the support of the RT model is and how uncertainty in it should be combined with uncertainty in the ELM.

Finally, the analysis in this thesis treated the pixel-based values of radiance as truth. In reality these are a product of the ATM sensor and the `azgcorr` software. They are subject to the limitations of engineering, varying environmental conditions and uncertainties associated with geometric correction and resampling (Curran & Hay 1986, Cracknell 1998, Anderson & Milton 2006). Future research should assess the likely impact of these considerations on the ELM.

9.5 Concluding statement

The research presented in this thesis began by investigating the typical approach to implementing the ELM. This was found to be problematic for reasons presented in section 5.1. The first part of the analysis sought to reduce uncertainty in the data input to the regression model and implemented the model on a per-pixel basis. By applying the simple

averaging block-pixel approach more effective use was made of the data and the uncertainty in the support was reduced by matching the supports of the reflectance and radiance data. This method was contingent on accurate positioning of the field data.

The regression model used in chapter 5 was inadequate because it yielded residuals that were heteroskedastic and spatially autocorrelated. By accounting for these, structural uncertainty in the model was reduced. This gave an improved understanding of uncertainty in the parameter estimates and model outputs.

Bibliography

Adams, J. (1995). *Risk*, Routledge, London.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. N. Petrov & F. Csaki (eds), *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, pp. 267–281.

Aman, A., Randriamanantena, H. P., Podaire, A. & Frouin, R. (1992). Upscale integration of Normalized Difference Vegetation Index - the problem of spatial heterogeneity, *IEEE Transactions on Geoscience and Remote Sensing* **30**(2): 326–338.

Anderson, K. (2005). *Temporal variability in calibration target reflectance: methods, models and applications*, Phd, University of Southampton.

Anderson, K. & Milton, E. J. (2005). *Characterisation of the apparent reflectance of a concrete calibration surface over different time scales*, Proceedings of the 9th International Symposium on Physical Measurements and Signatures in Remote Sensing (ISPMSRS), Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, China.

Anderson, K. & Milton, E. J. (2006). On the temporal stability of ground calibration targets: implications for the reproducibility of remote sensing methodologies, *International Journal of Remote Sensing* **27**(16): 3365–3374.

Armstrong, M. (1998). *Basic Linear Geostatistics*, Springer, London.

Ashton, E. A. (1998). Detection of subpixel anomalies in multispectral infrared imagery using an adaptive Bayesian classifier, *IEEE Transactions on Geoscience and Remote Sensing* **36**(2): 506–517.

Asrar, G. (ed.) (1989). *Theory and Application of Optical Remote Sensing*, Wiley, New York.

Atkinson, P. M. (1996). Simulating locational error in field–based measurements of reflectance, *in* A. Soares, J. Gomez-Hernandez & R. Froidevaux (eds), *geoENV I – Geostatistics for Environmental Applications*, Kluwer, London, pp. 297–308.

Atkinson, P. M. (1999a). Assessing accuracy in fuzzy land cover maps, *in* M. Barnsley & P. Pan (eds), *RSS99 - Earth Observation: From Data to Information*, Proceedings of the 25th Annual Conference of the Remote Sensing Society, Remote Sensing Society, University of Wales, UK, pp. 79–86.

Atkinson, P. M. (1999b). Geographical information science: geostatistics and uncertainty, *Progress in Physical Geography* **23**(1): 134–142.

Atkinson, P. M. (2000). Autologistic regression for flood zonation using SAR imagery, *in* P. Fisher (ed.), *RSS2000: Adding Value to Remotely Sensed Data*, Proceedings of the 26th Annual Conference of the Remote Sensing Society (CD–ROM), Remote Sensing Society, University of Leicester, UK. On CD-ROM.

Atkinson, P. M. (2005). Spatial prediction and surface modeling, *Geographical Analysis* **37**(2): 113–123.

Atkinson, P. M. & Curran, P. J. (1995). Defining an optimal size of support for remote-sensing investigations, *IEEE Transactions on Geoscience and Remote Sensing* **33**: 768–776.

Atkinson, P. M., Dunn, R. & Harrison, A. R. (1996). Measurement error in reflectance data and its implications for regularizing the variogram, *International Journal of Remote Sensing* **17**: 3735–3750.

Atkinson, P. M. & Emery, D. R. (1999). Exploring the relation between spatial structure and wavelength: implications for sampling reflectance in the field, *International Journal of Remote Sensing* **20**(13): 2663–2678.

Atkinson, P. M. & Foody, G. M. (2002). Uncertainty in remote sensing and GIS: An overview, *in* G. M. Foody & P. M. Atkinson (eds), *Uncertainty in Remote Sensing and GIS*, John Wiley, Chichester, pp. 1–35.

Atkinson, P. M. & Kelly, R. E. J. (1997). Scaling-up point snow depth data in the UK for comparison with SSM/I imagery, *International Journal of Remote Sensing* **18**(2): 437–443.

Atkinson, P. M. & Tate, N. J. (2000). Spatial scale problems and geostatistical solutions: A review, *Professional Geographer* **52**(4): 607–623.

Atkinson, P. M., Webster, R. & Curran, P. J. (1992). Cokriging with ground–based radiometry, *Remote Sensing of Environment* **41**(1): 45–60.

Atkinson, P. M., Webster, R. & Curran, P. J. (1994). Cokriging with airborne MSS imagery, *Remote Sensing of Environment* **50**(3): 335–345.

Augustin, N. H., Mugglestone, M. A. & Buckland, S. T. (1996). An autologistic model for the spatial distribution of wildlife, *Journal of Applied Ecology* **33**(2): 339–347.

Augustin, N. H., Mugglestone, M. A. & Buckland, S. T. (1998). The role of simulation in modelling spatially correlated data, *Environmetrics* **9**(2): 175–196.

Bamber, J. L., Vaughan, D. G. & Joughin, I. (2000). Widespread complex flow in the interior of the Antarctic ice sheet, *Science* **287**(5456): 1248–1250.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Monographs on Statistics and Applied Probability 101, Chapman and Hall / CRC, London.

Barford, N. C. (1985). *Experimental measurements: precision, error and truth*, second edn, John Wiley and Sons, Chichester.

Bates, P. D., Horritt, M., Smith, C. N. & Mason, D. (1997). Integrating remote sensing observations of flood hydrology and hydraulic modelling, *Hydrological Processes* **11**(14): 1777–1795.

Baveye, P. (2004). The emergence of a new kind of relativism in environmental modelling: a commentary, *Proceedings of the Royal Society of London Series A-Mathematical Physical And Engineering Sciences* **460**(2047): 2141–2146.

Ben-Dor, E., Kruse, F. A., Lefkoff, A. B. & Banin, A. (1994). Comparison of three calibration techniques for utilization of GER 63–channel aircraft scanner data of Makhtesh Ramon, Negev, Israel, *Photogrammetric Engineering and Remote Sensing* **60**(11): 1339–1354.

Ben-Dor, E. & Levin, N. (2000). Determination of surface reflectance from raw hyperspectral data without simultaneous ground data measurements: a case study of the GER 63-channel sensor data acquired over Naan, Israel, *International Journal of Remote Sensing* **21**(10): 2053–2074.

Berterretche, M., Hudak, A. T., Cohen, W. B., Maiersperger, T. K., Gower, S. T. & Dungan, J. (2005). Comparison of regression and geostatistical methods for mapping leaf area index (LAI) with Landsat ETM+ data over a boreal forest, *Remote Sensing of Environment* **96**(1): 49–61.

Best, N., Cockings, S., Bennett, J., Wakefield, J. & Elliott, P. (2001). Ecological regression analysis of environmental benzene exposure and childhood leukaemia: sensitivity to data inaccuracies, geographical scale and ecological bias, *Journal of the Royal Statistical Society Series A (Statistics in Society)* **164**: 155–174.

Bierkens, M. F. P., Fink, P. A. & Willingen, P. d. (2000). *Upscaling and Downscaling Methods for Environmental Research*, Kluwer Academic Publishers, London.

Blalock, H. M. (1964). *Causal inferences in non–experimental research*, University of California Press, Chapel Hill.

Box, G. E. P. & Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*, John Wiley and Sons, Chichester.

Brown, J. D. & Heuvelink, G. B. (2007). The data uncertainty engine (DUE): A software tool for assessing and simulating uncertain environmental variables, *Computers & Geosciences* **32**(1): 172–190.

Brus, D. J. & de Gruijter, D. J. (1997). Random sampling or geostatistical modelling? choosing between design–based sampling strategies soil (with discussion), *Geoderma* **80**(1): 1–44.

Buckland, S. T. & Elston, D. A. (1993). Empirical models for the spatial distribution of wildlife, *Journal of Applied Ecology* **30**: 478–495.

Burke, E. J., Banks, A. C. & Gurney, R. J. (1997). Remote sensing of soil-vegetation-atmosphere transfer processes, *Progress in Physical Geography* **21**(4): 549–572.

Burrough, P. A. & McDonnell, R. A. (1998). *Principles of Geographical Information Systems*, Oxford University Press, Oxford.

Campbell, J. B. (1996). *Introduction to Remote Sensing*, second edn, Taylor and Francis, London.

Carsel, R. F. & Parrish, R. S. (1988). Developing joint probability distributions of soil water retention characteristics, *Water Resources Research* **24**(5): 755–769.

Casella, G. & Berger, R. L. (1990). *Statistical Inference*, Duxbury Press, Belmont, California.

Caselles, V. & Lopez Garcia, M. J. (1989). An alternative simple approach to estimate atmospheric correction in multitemporal studies, *International Journal of Remote Sensing* **10**: 1127–1134.

Chambers, J. M., Cleveland, W. S., Kleiner, B. & Tukey, P. A. (1983). *Graphical Methods for Data Analysis*, Duxberry Press, Boston.

Chandrasekhar, S. (1950). *Radiative Transfer*, Dover, New York.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference, *Journal of the Royal Statistical Society Series A (Statistics in Society)* **158**: 419–466.

Chavez Jr., P. S. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data, *Remote Sensing of Environment* **24**(3): 459–479.

Chavez Jr., P. S. (1996). Image-based atmospheric corrections revisited and improved, *Photogrammetric Engineering And Remote Sensing* **62**(9): 1025–1036.

Chiles, J. P. (1976). How to adapt kriging to non–classical problems: three case studies, *in* M. Guarascio, M. David & C. Huijbregts (eds), *Avanced Geostatstics in the Mining Industry*, Reidel, Dordrecht, pp. 69–89.

Chiles, J.-P. & Delfiner, P. (1999). *Geostatistics: Modelling Spatial Uncertainty*, John Wiley and Sons, Chichester.

Christensen, O. F. & Ribeiro, P. J. (2002). geoRglm: A package for generalised linear spatial models, *R News* **2**(2): 26–28.

Clark, R. N., Swayze, G. A., King, T. V., Livo, K. E., Kokaly, R. F., Dalton, J. B., Vance, J. S., Rockwell, B. W. & McDougal, R. R. (2000). Surface reflectance calibration of terrestrial imaging spectroscopy data: a tutorial using AVIRIS, *Technical report*, USGS Spectroscopy Laboratory.

Clark, R. N., Swayze, G., Heidebrecht, K., Goetz, A. F. H. & Green, R. O. (1993). Comparison of methods for calibrating AVIRIS data to ground reflectance, *in* R. Green (ed.), *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop*, Vol. JPL Publication 93-26, Jet Propulsion Laboratory, Pasadena, pp. 35–36.

Cockings, S. & Martin, D. (2005). Zone design for environment and health studies using pre-aggregated data, *Social Science and Medicine* **60**: 2729–2742.

Collins Softback English Dictionary (1992). Harper Collins, Glasgow.

Cook, D. G. & Pocock, S. J. (1983). Multiple-regression in geographical mortality studies, with allowance for spatially correlated errors, *Biometrics* **39**(2): 361–371.

Cooley, T., Anderson, G. P., Felde, G. W., Hoke, M. L., Ratkowski, A. J., Chetwynd, J. H., Gardner, J. A., Adler-Golden, S. M., Matthew, M. W., Berk, A., Bernstein, L. S., Acharya, P. K., Miller, D. & Lewis, P. (2002). FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation, *IGARSS 2002: IEEE International Geoscience And Remote Sensing Symposium And 24th Canadian Symposium On Remote Sensing, Vols I-VI, Proceedings - Remote Sensing: Integrating Our View Of The Planet*, IEEE, New York, pp. 1414–1418.

Coops, N. & Culvenor, D. (2000). Utilizing local variance of simulated high spatial resolution imagery to predict spatial pattern of forest stands, *Remote Sensing of Environment* **71**(3): 248–260.

Cox, D. R. (2001). Another comment on the role of statistical methods, *British Medical Journal* **322**: 231.

Cracknell, A. P. (1998). Synergy in remote sensing – what's in a pixel, *International Journal of Remote Sensing* **19**(11): 2025–2047.

Crawley, M. J. & Harral, J. E. (2001). Scale dependence in plant biodiversity, *Science* **291**(5505): 864–868.

Cressie, N. (1990). The origins of kriging, *Mathematical Geology* **22**(3): 239–252.

Cressie, N. (1996). Change of support and the modifiable areal unit problem, *Geographical Systems* **3**(1): 159–180.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*, revised edn, John Wiley and Sons, New York.

Cressie, N. & Kornak, J. (2003). Spatial statistics in the presence of location error with an application to remote sensing of the environment, *Statistical Science* **18**(4): 436–456.

Cross, P. A., Barnes, J., Walker, A. H., Muller, J. P. & Morely, J. G. (2000). GPS validation of IfSAR digital elevation models from LANDMAP, *in* P. Fisher (ed.), *RSS2000: Adding Value to Remotely Sensed Data*, Proceedings of the 26th Annual Conference of the Remote Sensing Society (CD–ROM), Remote Sensing Society, University of Leicester, UK. On CD-ROM.

Curran, P. (1985). *Principles of Remote Sensing*, Longman, London.

Curran, P. J. (2002). Forword: Forward with uncertainty, *in* G. M. Foody & P. M. Atkinson (eds), *Uncertainty in Remote Sensing and GIS*, John Wiley, Chichester, pp. xi–xvi.

Curran, P. J. & Atkinson, P. M. (1998). Geostatistics and remote sensing, *Progress in Physical Geography* **22**: 61–78.

Curran, P. J., Atkinson, P. M., Foody, G. M. & Milton, E. J. (2000). Linking remote sensing, land cover and disease, *Advances in Parisitology* **47**: 37–80.

Curran, P. J. & Hay, A. M. (1986). The importance of measurement error for certain procedures in remote sensing at optical wavelengths, *Photogrammetric Engineering and Remote Sensing* **52**(2): 229–241.

Curran, P. J. & Milton, E. J. (1983). The relationships between the chlorophyll concentration, LAI and reflectance of a simple vegetation canopy, *International Journal of Remote Sensing* **4**(2): 247–255.

Curran, P. J. & Williamson, H. D. (1985). The accuracy of ground data used in remote-sensing investigations, *International Journal of Remote Sensing* **6**: 1637–1651.

Danson, F. M. & Plummer, S. E. (eds) (1995). *Advances in Environmental Remote Sensing*, John Wiley and Sons, Chichester.

Davis, M. W. (1987a). Generating large stochastic simulations - the matrix polynomial-approximation method, *Mathematical Geology* **19**(2): 99–107.

Davis, M. W. (1987b). Production of conditional simulations via the LU triangular decomposition of the covariance-matrix, *Mathematical Geology* **19**(2): 91–98.

DeGroot, M. H. & Schervish, M. J. (2002). *Probability and Statistics*, Addison-Wesley, London.

Deschamps, P. Y., Herman, M. & Tanre, D. (1983). Modeling of the atmospheric effects and its application to the remote sensing of ocean color, *Applied Optics* **22**(23): 3751–3758.

Deutsch, C. V. & Journel, A. G. (1998). *GSLIB: Geostatistical Software Library and User's Guide*, second edn, Oxford University Press, New York.

Diggle, P. J. & Ribeiro Jr., P. (2002). Bayesian inference in model–based geostatistics, *Geographical and Environmental Modelling* **6**(2): 129–146.

Diggle, P. J. & Ribeiro Jr., P. J. (2006). *Model-based Geostatistics*, Springer Verlag, New York.

Diggle, P. J., Ribeiro Jr., P. J. & Christensen, O. F. (2003). An introduction to model–based geostatistics, *in* J. Møller (ed.), *Spatial Statistics and Computational Methods*, Springer, London, pp. 43–86.

Diggle, P. J., Tawn, J. A. & Moyeed, R. A. (1998). Model–based geostatistics, *Applied Statistics* **47**(3): 299–350.

Dobson, A. J. (1990). *An Introduction to Generalized Linear Models*, Chapman and Hall/CRC, London.

Dorren, L. K. A. & Heuvelink, G. B. M. (2004). Effect of support size on the accuracy of a distributed rockfall model, *International Journal of Geographical Information Science* **18**(6): 595–609.

Draper, D. (1995). Assessment and propagation of model uncertainty, *Journal of the Royal Statistical Society Series B (Methodological)* **57**(1): 45–97.

Duggin, M. J. & Philpson, W. R. (1982). Field measurement of reflectance: some major considerations, *Applied Optics* **21**(15): 2833–2840.

Dungan, J. (1998). Spatial prediction of vegetation quantities using ground and image data, *International Journal of Remote Sensing* **19**(2): 267–285.

Dungan, J. L. (1999). Conditional simulation: An alternative for achieving mapping objectives, *in* A. Stein, F. van der Meer & B. Gorte (eds), *Spatial Statistics for Remote Sensing*, Kluwer, London, pp. 135–152.

Dungan, J. L. (2002). Towards a comprehensive view of uncertainty in remote sensing analysis, *in* G. M. Foody & P. M. Atkinson (eds), *Uncertainty in Remote Sensing and GIS*, John Wiley, Chichester, pp. 25–35.

Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M. J., Jakomulska, A., Miriti, M. & Rosenberg, M. S. (2002). A balanced view of scale in spatial statistical analysis, *Ecography* **25**(5): 626–640.

Dwyer, J. L., Kruse, F. A. & Lefkoff, A. B. (1995). Effects of empirical versus model–based reflectance on automated analysis of imaging spectrometer data: A case study from the Drum Mountains, Utah, *Photogrammetric Engineering and Remote Sensing* **61**(10): 1247–1254.

Elith, J., Burgman, M. A. & Regan, H. M. (2002). Mapping epistemic uncertainties and vague concepts in predictions of species distribution, *Ecological Modelling* **157**(2-3): 313–329.

ESRI (2004). *Using ArcGIS Geostatistical Analyst: ArcGIS 9*, Environmental Systems Research Institute, Redlands, CA.

Farrand, W. H., Singer, R. B. & Merényi, E. (1994). Retrieval of apparent surface reflectance from AVIRIS data: A comparison of empirical line, radiative transfer and spectral mixture methods, *Remote Sensing of Environment* **47**(1): 311–321.

Fernandes, R. & Leblanc, S. G. (2005). Parametric (modified least squares) and non-parametric (Theil-Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors, *Remote Sensing of Environment* **95**(3): 303–316.

Ferrier, G. (1995). Evaluation of apparent surface reflectance estimation methodologies, *International Journal of Remote Sensing* **17**(12): 2291–2297.

Ferrier, G. & Wadge, G. (1996). The application of imaging spectrometry data to mapping alteration zones associated with gold mineralization in southern Spain, *International Journal of Remote Sensing* **16**(2): 331–350.

Ferson, S., Nelsen, R. B., Hajagos, J., Berleant, D. J., Zhang, J., Tucker, W. T., Ginzburg, L. & Oberkampf, W. L. (2004). Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis, *Technical Report SAND2004-3072*, Sandia National Laboaratories.

Feynman, R. P. (1988). *What do you Care What Other People Think?*, Harper Collins, London.

Fisher, P. F. (1999). Models of uncertainty in spatial data, *in* P. A. Longley, M. F. Goodchild, D. J. Maguire & D. W. Rhind (eds), *Geographical Information Systems: Principles, Techniques, Management, and Applications*, John Wiley and Sons, Chichester, pp. 190–206.

Fisher, P. F. & Langford, M. (1995). Modeling the errors in areal interpolation between zonal systems by Monte-Carlo simulation, *Environment and Planning A* **27**(2): 211–224.

Foody, G. M. (2003). Uncertainty, knowledge discovery and data mining in GIS, *Progress In Physical Geography* **27**(1): 113–121.

Foody, G. M. & Atkinson, P. M. (eds) (2002). *Uncertainty in Remote Sensing and GIS*, John Wiley, Chichester.

Foody, G. M. & Curran, P. J. (1994a). Scale and environmental remote sensing, *in* G. M. Foody & P. J. Curran (eds), *Environmental Remote Sensing from Regional to Global Scales*, John Wiley and Sons, Chichester, pp. 223–232.

Foody, G. M. & Curran, P. J. (eds) (1994b). *Environmental Remote Sensing from Regional to Global Scales*, John Wiley and Sons, Chichester.

Forster, R. M. & Jesus, B. (2006). Field spectroscopy of estuarine intertidal habitats, *International Journal of Remote Sensing* **27**(17): 3657–3669.

Fotheringham, A. S., Brunsdon, C. & Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*, Sage Publications, London.

Fotheringham, A. S., Brunsdon, C. & Charlton, M. (2002). *Geographically Weighted Regression: the analysis of spatially varying relationships*, John Wiley and Sons, Chichester.

Fotheringham, A. S. & Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical-analysis, *Environment and Planning A* **23**(7): 1025–1044.

Freemantle, J. R., Pu, R. & Miller, J. R. (1992). Calibration of imaging spectrometer data to reflectance using pseudo-invariant features, *in* J. K. Hornsby, D. J. King & N. A. Prout (eds), *Proceedings of the 15th Canadian Symposium on Remote Sensing*, Ontario Centre for Remote Sensing, Toronto, pp. 452–457.

Fuller, W. A. (1987). *Measurement Error Models*, John Wiley and Sons, Chichester.

Gabrosek, J. & Cressie, N. (2001). The effect on attribute prediction of locational uncertainty in spatial data, *Technical report 674*, Department of Statistics, Ohio State University.

Gabrosek, J. & Cressie, N. (2002). The effect on attribute prediction of location uncertainty in spatial data, *Geographical Analysis* **34**(3): 262–284.

Gabrosek, J. G. (1999). *The effect of locational uncertainty in geostatististics*, PhD thesis, Iowa Sate University.

Gao, B. C. & Goetz, A. F. H. (1990). Column atmospheric water-vapor and vegetation liquid water retrievals from airborne imaging spectrometer data, *Journal of Geophysical Research* **95**(D4): 3549–3564.

Gao, J. (2006). Quantification of grassland properties: how it can benefit from geoinformatic technologies?, *International Journal of Remote Sensing* **27**(7): 1351–1365.

Gehlke, C. E. & Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material, *Journal American Statisitical Association* **29**(185): 169–170.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (1995). *Bayesian Data Analysis*, Chapman and Hall, London.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2004). *Bayesian Data Analysis*, second edn, Chapman and Hall, London.

Gilks, W., Richardson, S. & Spiegelhalter, D. (eds) (1996). *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Gilmour, A. R., Thompson, R. & Cullis, B. R. (1995). Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models, *Biometrics* **51**(4): 1440–1450.

Goel, N. S. (1989). Inversion of canopy reflection models for estimation of biophysical parameters from reflectance data, *in* G. Asrar (ed.), *Theory and Application of Optical Remote Sensing*, Wiley, New York, pp. 205–251.

Goetz, A. F. H., Evans, K. F. & Ferri, M. (2003). Removing thin cirrus cloud effects in Hyperion data using the 1.38 and 1.87 μm water vapor absorption bands, *IGARSS 2003: IEEE International Geoscience And Remote Sensing Symposium, Vols I - VII, Proceedings - Learning From Earth's Shapes And Sizes*, IEEE, New York, pp. 83–85.

Goetz, A. F. H., Ferri, M., Kindel, B. & Qu, Z. (2002). Atmospheric correction of Hyperion data and techniques for dynamic scene correction, *IGARSS 2002: IEEE International Geoscience And Remote Sensing Symposium And 24th Canadian Symposium On Remote Sensing, Vols I-VI, Proceedings - Remote Sensing: Integrating Our View Of The Planet*, IEEE, New York, pp. 1408–1410.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*, Oxford University Press, Oxford.

Goovaerts, P. (1999). Using elevation to aid the geostatistical mapping of rainfall erosivity, *Catena* **34**(3-4): 227–242.

Gotway, C. A. & Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction, *Journal of Agricultural, Biological and Environmental Statistics* **2**(2): 157–178.

Gotway, C. A. & Young, L. J. (2002). Combining incompatible spatial data, *Journal of the American Statistical Association* **97**(458): 632–648.

Gu, X. F., Guyot, G. & Verbrugghe, M. (1992). Evaluation of measurement errors in ground surface reflectance for satellite calibration, *International Journal of Remote Sensing* **13**(14): 2531–2546.

Hamm, N., Atkinson, P. M. & Milton, E. J. (2004). On the effect of positional uncertainty in field measurements on the atmospheric correction of remotely sensed imagery, *in* X. Sanchez-Vila, J. Carrera & J. J. Gomez-Hernandez (eds), *geoENV IV –*

Geostatistics for Environmental Applications, Kluwer Academic Publishers, London, pp. 91–102.

Harmon, R. & Challenor, P. (1997). A Markov chain Monte Carlo method for estimation and assimilation into models, *Ecological Modelling* **101**(1): 41–59.

Harris, T. R. & Johnson, D. E. (1996). A regression model with spatially correlated errors for comapring remote sensing and in-situ measurements of a grassland site, *Journal of Agricultural, Biological and Environmental Statistics* **1**(2): 190–204.

Helton, J. C., Johnson, J. D. & Oberkampf, W. L. (2004). An exploration of alternative approaches to the representation of uncertainty in model predictions, *Reliability Engineering and System Safety* **85**(1-3): 39–71.

Hengl, T., Heuvelink, G. B. M. & Stein, A. (2003). Comparison of kriging with external drift and regression-kriging, *Technical note*, ITC.

Hengl, T., Heuvelink, G. B. M. & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging, *Geoderma* **120**(1-2): 75–93.

Hess, M., Koepke, P. & Schult, I. (1998). Optical properties of aerosols and clouds: The software package OPAC, *Bulletin of the American Meteorological Society* **79**(5): 831–844.

Heuvelink, G. B. M. (1998). *Error Propagation in Environmental Modelling*, Taylor and Francis, London.

Heuvelink, G. B. M. (2002). Analysing uncertainty propagation in GIS: Why it is not that simple, *in* G. M. Foody & P. M. Atkinson (eds), *Uncertainty in Remote Sensing and GIS*, John Wiley, Chichester, pp. 155–165.

Heuvelink, G. B. M. & Pebesma, E. (1999). Spatial aggregation and soil process modelling, *Geoderma* **89**(1–2): 47–65.

Heywood, I., Cornelius, S. & Carver, S. (2006). *An Introduction to Geographical Information Systems*, third edn, Prentice Hall, Harlow.

Holton, G. (2004). Defining risk, *Financial Analysts Journal* **60**(6): 19–25.

Isaaks, E. H. & Srivistava, R. M. (1989). *An Introduction to Applied Geostatistics*, Oxford University Press, Oxford.

Jackson, C., Best, N. & Richardson, S. (2006). Improving ecological inference using individual-level data, *Statistics In Medicine* **25**(12): 2136–2159.

Jago, R. A., Cutler, M. E. J. & Curran, P. J. (1999). Estimating canopy chlorophyll concentration from field and airborne spectra, *Remote Sensing of Environment* **68**(3): 217–224.

Jelinski, D. E. & Wu, J. G. (1996). The modifiable areal unit problem and implications for landscape ecology, *Landscape Ecology* **11**(3): 129–140.

Johnson, M. E. (1987). *Multivariate Statistical Simulation*, Wiley, New York.

Journel, A. G. (1973). Geostatistics for conditional simulation of ore bodies, *Economic Geology* **69**: 527–545.

Journel, A. G. (1996a). Conditional simulation of geologically averaged block permeabilities, *Journal of Hydrology* **183**(1–2): 23–35.

Journel, A. G. (1996b). Modelling uncertainty and spatial dependence: Stochastic imaging, *International Journal of Geographical Information Systems* **10**(5): 517–522.

Journel, A. G. & Huijbregts, J. (1978). *Mining Geostatistics*, Academic Press, London.

Justice, C. & Townshend, J. (1981). Integrating ground data with remote sensing, *in* J. Townshend (ed.), *Terrain Analysis and Remote Sensing*, George Allen and Unwin, London, pp. 38–58.

Karpouzli, E. & Malthus, T. (2003). The empirical line method for the atmospheric correction of IKONOS imagery, *International Journal of Remote Sensing* **24**(5): 1143–1150.

Kaufman, Y. J. (1989). The atmospheric effect on remote sensing and its correction, *in* G. Asrar (ed.), *Theory and Application of Optical Remote Sensing*, Wiley, New York, pp. 336–348.

Kaufman, Y. J., Wald, A. E., Remer, L. A., Gao, B. C., Li, R. R. & Flynn, L. (1997). The MODIS 2.1-μm channel - correlation with visible reflectance for use in remote sensing of aerosol, *IEEE Transactions on Geoscience and Remote Sensing* **35**(5): 1286–1298.

Kendall, M. G. & Stuart, A. (1967). *The Advanced Theory of Statistics*, Vol. 2: Inference and Relationships, second edn, Charles Griffin and Co. Ltd, London.

Kermack, K. & Haldane, J. (1950). Organic correlation and allometry, *Biometrika* **37**(1/2): 40–41.

Kimes, D. S., Knyazikhin, Y., Privette, J. L., Abuelgasim, A. A. & Gao, F. (2000). Inversion methods for physically–based models, *Remote Sensing Reviews* **18**: 381–439.

Kitmitto, K., Muller, J. P., Morley, J. G., Walker, A., Mitchell, K., Barnes, J., Cross, P. A., Dowman, I. J., Smith, A., Chugani, K. & Quarmby, N. (2000). LANDMAP – a project for mapping the British Isles from space: Multi–sensor geocoded images based on based on ERS-tandem SAR interferometry, *Proceedings of GIS Research UK 2000*, GISRUK, York.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*, Houghton Mifflin, Boston.

Knotters, M., Brus, D. J. & Voshaar, J. H. O. (1995). A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations, *Geoderma* **67**(3-4): 227–246.

Kruse, F. A., Kierein-Young, K. S. & Boardman, J. W. (1990). Mineral mapping at Cuprite, Nevada with a 63–channel imaging spectrometer, *Photogrammetric Engineering and Remote Sensing* **56**(1): 83–92.

Kuha, J. (2004). AIC and BIC - comparisons of assumptions and performance, *Sociological Methods and Research* **33**(2): 188–229.

Lark, R. M. (2000). Regression analysis with spatially autocorrelated error: simulation studies and application to mapping of soil organic matter, *International Journal of Geographical Information Science* **14**(3): 247–264.

Lark, R. M. & Cullis, B. R. (2004). Model-based analysis using REML for inference from systematically sampled data on soil, *European Journal of Soil Science* **55**(4): 799–813.

Lark, R. M. & Webster, R. (2006). Geostatistical mapping of geomorphic variables in the presence of trend, *Earth Surface Processes and Landforms* **31**(7): 862–874.

Larsson, H. (1993). Linear regression for canopy cover estimation in Acacia woodlands using Landsat-TM, -MSS and SPOT HRV XS data, *International Journal of Remote Sensing* **14**(11): 2129–2136.

Lawless, K. P., Milton, E. J. & Anger, C. M. (1998). Investigation of changes in the reflectance of ground calibration targets (asphalt and concrete), *Information for Sustainability*, 27th International Symposium on Remote Sensing of Environment, ERIM, Ann Arbour, Michigan, pp. 597–600.

Letham, L. (2003). *GPS Made Easy: Using Global Positioning Systems in the Outdoors*, Cordee, Leicester.

Lewis, P., Disney, M. I. & Riedmann, M. (1999). Application of the botanical plant modelling system (BPMS) to the analysis of spatial information in remotely sensed imagery, *in* M. Barnsley & P. Pan (eds), *RSS99 - Earth Observation: From Data to Information*, Proceedings of the 25th Annual Conference of the Remote Sensing Society, Remote Sensing Society, University of Wales, UK, pp. 507–514.

Li, J. L., Liang, T. G. & Chen, Q. G. (1998). Estimating grassland yields using remote sensing and GIS technologies in China, *New Zealand Journal of Agricultural Research* **41**(1): 31–38.

Liang, S. (2004). *Quantitative Remote Sensing of Land Surfaces*, Hoboken, New Jersey, John Wiley and Sons.

Liang, S. L., Fallah-Adl, H., Kalluri, S., JaJa, J., Kaufman, Y. J. & Townshend, J. R. G. (1997). An operational atmospheric correction algorithm for Landsat Thematic

Mapper imagery over the land, *Journal of Geophysical Research* **102**(D14): 17173–17186.

Liang, S. L., Fang, H. & Chen, M. (2001). Atmospheric correction of Landsat ETM+ land surface imagery - Part I: Methods, *IEEE Transactions on Geoscience and Remote Sensing* **39**(11): 2490–2498.

Liang, S., Strahler, A. H. & Walthall, C. L. (1999). Retrieval of land surface albedo from satellite observations: A simulation study, *Journal of Applied Meteorology* **38**(6): 712–725.

Lillesand, T. M. & Kiefer, R. W. (1994). *Remote Sensing and Image Interpretation*, third edn, John Wiley and Sons, New York.

Lillesand, T. M., Kiefer, R. W. & Chipman., J. W. (2004). *Remote Sensing and Image Interpretation*, fifth edn, John Wiley and Sons, Hoboken, NJ.

Llewellyn, G. M., Kooistra, L. & Curran, P. J. (2000). The effect of soil contamination on grassland spectra, *in* P. Fisher (ed.), *RSS2000: Adding Value to Remotely Sensed Data*, Proceedings of the 26th Annual Conference of the Remote Sensing Society (CD–ROM), Remote Sensing Society, University of Leicester, UK. On CD-ROM.

Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (2005). *Geographic Information Systems and Science*, second edn, John Wiley and Sons, London.

Maling, D. H. (1989). *Measurements from Maps*, Pergamon Press, New York.

Maracci, G., Tassone, G. & Toselli, F. (1981). The use of a reference panel in the visible and thermal infrared as a tool for calibrating airborne multispectral scanners, *in* G. Guyot & M. Verbrugghe (eds), *Spectral Signatures of Objects in Remote Sensing*, International Society for Photogrammetry and Remote Sensing, CNES, Avignon, France, pp. 165–172.

Mardia, K. V. & Marshall, R. J. (1984). Maximum-likelihood estimation of models for residual covariance in spatial regression, *Biometrika* **71**(1): 135–146.

Martin, D. (1996). *Geographic Information Systems: Socioeconomic Applications*, Routledge, London.

Martin, D. & Atkinson, P. (2001). Investigating the spatial linkage of primary school performance and catchment characteristics, *Geographical and Environmental Modelling* **5**(1): 67–83.

Martonchik, J. V., Bruegge, C. J. & Strahler, A. H. (2000). A review of reflectance nomenclature used in remote sensing, *Remote Sensing Reviews* **19**(1): 9–20.

Mátern, B. (1986). *Spatial Variation*, Springer Verlag, Berlin.

Mather, P. M. (1999). *Computer Processing of Remotely-Sensed Images: An Introduction*, second edn, John Wiley and Sons, Chichester.

Matheron, G. (1963). Principles of geostatistics, *Economic Geology* **58**: 1246–1266.

Matheron, G. (1965). *Les variables regionalisees et leur estimation*, Masson, Paris.

May, R. (2001). Risk and uncertainty, *Nature* **411**(6840): 891–891.

McArdle, S. S., Miller, J. R. & Freemantle, J. R. (1992). Airborne image acquisition under cloud: Preliminary comparisons with clear-sky scene radiance and reflectance imagery, *in* J. K. Hornsby, D. J. King & N. A. Prout (eds), *Proceedings of the 15th Canadian Symposium on Remote Sensing*, Ontario Centre for Remote Sensing, Toronto, pp. 446–451.

Mie, G. (1908). A contribution to the optics of turbid media, especially colloidal metallic suspensions (in German), *Annals Physics (Leipzig)* **25**: 377.

Milton, E. J. (1980). A portable multiband radiometer for ground data collection in remote sensing, *International Journal of Remote Sensing* **1**(2): 153–165.

Milton, E. J. (1981a). Does the use of two radiometers correct for irradiance changes during measurements?, *Photogrammetric Engineering and Remote Sensing* **47**(8): 1223–1225.

Milton, E. J. (1981b). Empirical studies of the reflectance of complex natural ssurfaces using the Milton Multiband Radiometer, *in* G. Guyot & M. Verbrugghe (eds), *Spectral Signatures of Objects in Remote Sensing*, International Society for Photogrammetry and Remote Sensing, CNES, Avignon, France, pp. 173–181.

Milton, E. J. (1982). Field measurements of reflectance factors: A further note, *Photogrammetric Engineering and Remote Sensing* **48**(9): 1474–1476.

Milton, E. J. (1987). Principles of field spectroscopy, *International Journal of Remote Sensing* **8**(12): 1807–1827.

Milton, E. J. (2001). User notes for the Milton Multiband Radiometer, *Technical report v 6.0*, Department of Geography, University of Southampton, UK.

Milton, E. J., Fox, N. & Mackin, S. (2004). Calibration, validation and the NERC Airborne Remote Sensing Facility, *Annual Conference of the Remote Sensing and Photogrammetry Society*, Remote Sensing and Photogrammetry Society, Aberdeen, Scotland. On CD-ROM.

Milton, E. J. & Goetz, A. F. H. (1997). Atmospheric influences on field spectrometry: Observed relationships between spectral irradiance and the variance in spectral reflectance, *in* G. Guyot & T. Phulpin (eds), *Seventh International Symposium on Physical Measurements and Signatures in Remote Sensing*, International Society for

Photogrammetry and Remote Sensing, Balkema, Rotterdam, Val D'Isére, France, pp. 109–114.

Milton, E. J., Lawless, K. P., Roberts, A. & Franklin, S. E. (1997). The effect of unresolved scene elements on the spectral response of calibraton targets: An example, *Canadian Journal of Remote Sensing* **23**(3): 252–256.

Milton, E. J., Rollin, E. M. & Brown, K. (2000). Practical methodologies for the reflectance calibration of *casi* data (activities of the NERC Equipment Pool for Field Spectroscopy (EPFS) in support of the NERC Airborne Remote Sensing Facility), *NERC ARSF Meeting*, BGS Keyworth.

Milton, E. J., Rollin, E. M. & Emery, D. R. (1995). Advances in field spectroscopy, *in* F. M. Danson & S. E. Plummer (eds), *Advances in Environmental Remote Sensing*, John Wiley and Sons, Chichester, pp. 9–32.

Milton, E. J., Smith, G. M. & Lawless, K. P. (1996). Preparatory research to develop an operational method to calibrate airborne sensor data using a network of ground calibration sites, *Proceedings of the Second International Airborne Remote Sensing Conference and Exhibition*, ERIM, Ann Arbour, Michigan, pp. 132–141.

Møller, J. (ed.) (2003). *Spatial Statistics and Computational Methods*, Springer–Verlag, London.

Montello, D. R. (2001). Scale in geography, *in* N. J. Smelser & P. B. Bates (eds), *International Encyclopedia of the Social and Behavioral Sciences*, Permagon Press, Oxford, pp. 13501–13504.

Moran, M. S., Bryant, R., Holifield, C. D. & McElroy, S. (2003). Refined empirical line approach for retrieving surface reflectance from EO-1 ALI images, *IEEE Transactions on Geoscience and Remote Sensing* **41**(6): 1411–1414.

Moran, M. S., Bryant, R., Thome, K., Ni, W., Nouvellon, Y., Gonzalez-Dugo, M. P. & Qi, J. (2001). A refined empirical line approach for reflectance factor retrieval from Landsat-5 TM and Landsat-7 ETM+, *Remote Sensing of Environment* **78**(1-2): 71–82.

Moran, M. S., Jackson, R. D., Slater, P. N. & Teillet, P. M. (1992). Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output, *Remote Sensing of Environment* **41**(2-3): 169–184.

Mowrer, H. T. & Congalton, R. G. (2000). *Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing*, Ann Arbor, Chelsea, MI.

Muller, J. P., Morley, J. G., Walker, A., Barnes, J., Cross, P. A., Dowman, I. J., Mitchell, K., Smith, A., Chugani, K. & Kitmitto, K. (1999). The LANDMAP project for

the creation of multi-sensor geocoded and topographic map products for the British Isles based on ERS-tandem interferometry, *Advancing ERS SAR Interferometry from Applications towards Operations, Proceedings of the Second International Workshop on ERS SAR Interferometry*, ESA, Belgium.

Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wasserman, W. (1996). *Applied Linear Statistical Models*, fourth edn, Irwin, London.

Ni, W., Woodcock, C. E. & Jupp, D. L. B. (1999). Variance in bidirectional reflectance over discontinuous plant canopies, *Remote Sensing of Environment* **69**(1): 1–15.

Nicodemus, F. W., Richmond, J. C., Hsia, J. J., Ginsberg, I. W. & Limperis, T. (1977). Geometrical considerations for nomenclature for reflectance, *Technical report*, Department of Commerce, National Bureau of Standards, USA.

Oberkampf, W. L., Helton, J. C., Joslyn, C. A., Wojtkiewicz, S. F. & Ferson, S. (2004). Challenge problems: uncertainty in system response given uncertain parameters, *Reliability Engineering and System Safety* **85**(1-3): 11–19.

Odeh, I. O. A., Mcbratney, A. B. & Chittleborough, D. J. (1995). Further results on prediction of soil properties from terrain attributes - heterotopic cokriging and regression-kriging, *Geoderma* **67**(3-4): 215–226.

O'Hagan, A. & Oakley, J. E. (2004). Probability is perfect, but we can't elicit it perfectly, *Reliability Engineering and System Safety* **85**(1-3): 239–248.

Olea, R. A. (1975). Optimum mapping techniques using regionalized variable theory, *Report 2*, Kansas University Geological Survey Series on Spatial Analysis, University of Kansas, Lawrence.

Oliver, M. A. & Webster, R. (1986). Semi-variograms for modelling the spatial pattern of landform and soil properties, *Earth Surface Processes and Landforms* **11**: 491–504.

Oliver, M., Webster, R. & Gerrard, J. (1989). Geostatistics in physical geography. Part I: Theory, *Transactions of the Institute of British Geographers* **14**(3): 259–269.

Omre, H. (1987). Bayesian kriging – merging observations and qualified guesses in kriging, *Mathematical Geology* **19**(1): 25–39.

Openshaw, S. & Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem, *in* N. Wrigley (ed.), *Statistical Applications in the Spatial Sciences*, Pion Ltd, London, pp. 127–144.

Ordnance Survey (1999). A guide to coordinate systems in Great Britain: An introduction to mapping coordinate systems and the use of GPS datasets with Ordnance Survey mapping, *Technical report 1.1*, Department of Geodesy, Ordnance Survey, UK.

Ordnance Survey (2001a). Land–Form PANORAMA user guide, *Technical report 3.0*, Department of Geodesy, Ordnance Survey, UK.

Ordnance Survey (2001b). Land–Form PROFILE user guide, *Technical report 4.0*, Department of Geodesy, Ordnance Survey, UK.

Ordnance Survey (2002a). A guide to coordinate systems in Great Britain: An introduction to mapping coordinate systems and the use of GPS datasets with Ordnance Survey mapping, *Technical report 1.4*, Department of Geodesy, Ordnance Survey, UK.

Ordnance Survey (2002b). Land–line user guide, *Technical report 4.0*, Department of Geodesy, Ordnance Survey, UK.

Oudemans, P. V., Pozdnyakova, L., Hughes, M. G. & Rahman, F. (2002). GIS and remote sensing for detecting yield loss in cranberry culture, *Journal of Nematology* **34**(3): 207–212.

Paige, G. B. & Hillel, D. (1993). Comparison of three methods for assessing soil hydraulic properties, *Soil Science* **155**(3): 174–189.

Pardo-Igúzquiza, E., Chica-Olmo, M. & Atkinson, P. M. (2006). Downscaling cokriging for image sharpening, *Remote Sensing of Environment* **102**(1-2): 86–98.

Pebesma, E. J. (2003). Gstat: multivariable geostatistics for S, *The Third International Workshop on Distributed Statistical Computing (DSC2003)*, Universitat Wien, Vienna, Austria.

Pebesma, E. J. & Wesseling, C. G. (1998). Gstat, a program for geostatistical modelling, prediction and simulation, *Computers & Geosciences* **24**(1): 17–31.

Peddle, D. R., White, H. P., Soffer, R. J., Miller, J. R. & LeDrew, E. F. (2001). Reflectance processing of remote sensing spectroradiometer data, *Computers & Geosciences* **27**(2): 203–213.

Perry, E. M., Warner, T. & Foote, P. (2000). Comparison of atmospheric modelling using empirical line fitting for mosaicking HYDICE imagery, *International Journal of Remote Sensing* **21**(4): 799–803.

Philipson, W. R., Gordon, D. K., Philpot, W. D. & Duggin, M. J. (1989). Field reflectance calibration with grey standard reflectors, *International Journal of Remote Sensing* **10**(6): 1035–1039.

Prince, S. D. & Astle, W. L. (1986). Satellite remote-sensing of rangelands in Botswana 1. Landsat MSS and herbaceous vegetation, *International Journal of Remote Sensing* **7**(11): 1533–1553.

Raffy, M. (1994). Change of scale theory - a capital challenge for space observation of earth - introduction, *International Journal of Remote Sensing* **15**(12): 2353–2357.

Raffy, M. & Blamont, D. (2003). Radiometric nomenclatures and spatial resolutions, *International Journal of Remote Sensing* **24**(11): 2289–2307.

Regan, H. M., Colyvan, M. & Burgman, M. A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology, *Ecological Applications* **12**(2): 618–628.

Reynolds, W. D. (1994). Comments on "Comparison of three methods for assessing soil hydraulic properties" by G.B. Paige and D. Hillel, *Soil Science* **157**(2): 116–119.

Reynolds, W. D. & Zebchuk, W. D. (1996). Hydraulic conductivity in a clay soil: two measurement techniques and spatial characterisation, *Soil Science Society of America Journal* **60**: 1679–1685.

Ribeiro Jr., P. J., Christensen, O. F. & Diggle, P. J. (2003). Geostatistical software – geoR and geoRglm, *The Third International Workshop on Distributed Statistical Computing (DSC2003)*, Universitat Wien, Vienna, Austria.

Ribeiro, P. J. & Diggle, P. J. (2001). geoR : A package for geostatistical analysis, *R News* **1**(2): 14–18.

Richards, J. A. & Jia, X. (1999). *Remote Sensing Digital Image Analysis: An Introduction*, third edn, Springer, New York.

Roberts, D. A., Yamaguchi, Y. & Lyon, R. J. P. (1986). Comparison of various methods for calibration of AIS data, *Second Airborne Imaging Spectrometer Data Analysis Workshop*, JPL Publication 86–35, Jet Propulsion Laboratory, Jet Propulsion Laboratory, Pasadena, CA, pp. 21–30.

Robinson, A. H., Morrison, J. L., Muehrcke, P. C., Kimerling, A. J. & Guptill, S. C. (1995). *Elements of Cartography*, John Wiley and Sons, Chichester.

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding: Methods for Atmospheric Sounding: Theory and Practice*, World Scientific, London.

Rollin, E. M., Milton, E. J. & Emery, D. R. (2000). Reference panel anisotropy and diffuse radiation – some implications for field spectroscopy, *International Journal of Remote Sensing* **21**(15): 2799–2810.

Roujean, J. L., Leroy, M. & Deschamps, P. Y. (1992). A bidirectional reflectance model of the earth's surface for the correction of remote sensing data, *Journal of Geophysical Research* **97**: 20455–20468.

Roy, D. P., Devereux, B., Grainger, B. & White, S. J. (1997). Parameteric geometric correction of airborne thematic mapper imagery, *International Journal of Remote Sensing* **18**(9): 1865–1887.

Rufino, G., Moccacia, A. & Esposito, S. (1998). DEM generation by means of ERS tandem data, *IEEE Transactions on Geoscience and Remote Sensing* **36**(6): 1905–1912.

Sadahiro, Y. (2000a). Accuracy of count data estimated by the point-in-polygon method, *Geographical Analysis* **32**(1): 64–89.

Sadahiro, Y. (2000b). Accuracy of count data transferred through the areal weighting interpolation method, *International Journal of Geographical Information Science* **14**(1): 25–50.

Salvador, R. (1999). A parametric model for estimating relations between unprecisely located field measurements and remotely sensed data, *Remote Sensing of Environment* **67**(1): 99–107.

Salvador, R. & Pons, X. (1998). On the reliability of Landsat TM for estimating forest variables by regression techniques: A methodological analysis, *IEEE Transactions on Geoscience and Remote Sensing* **36**(6): 1888–1897.

Salvatori, V., Atkinson, P. M., Milton, E. J. & Emery, D. R. (2000). Wavelength–dependent spatial variation in the reflectance of 'homogeneous' ground calibration targets, *International Archives of Photogrammetry and Remote Sensing* **33**(1): 192–99.

Salvatori, V., Skidmore, A. K., Corsi, F. & van der Meer, F. (1999). Estimating temporal independence of radio-telemetry data on animal activity, *Journal of Theoretical Biology* **198**(4): 567–574.

Sanchez-Vila, X., Carrera, J. & Gomez-Hernandez, J. J. (eds) (2004). *geoENV IV – Geostatistics for Environmental Applications*, Kluwer Academic Publishers, London.

Sandmeier, S., Muller, C., Hosgood, B. & Andreoli, G. (1998). Sensitivity analysis and quality assessment of laboratory BRDF data, *Remote Sensing of Environment* **64**(2): 176–191.

Sandmeier, S. R. (2000). Acquisition of bidirectional reflectance factor data with field goniometers, *Remote Sensing of Environment* **73**(3): 257–269.

Sandmeier, S. R. & Itten, K. I. (1999). A field goniometer system (FIGOS) for acquisition of hyperspectral BRDF data, *IEEE Transactions on Geoscience and Remote Sensing* **37**(2): 978–986.

Schaepman-Strub, G., Schaepman, M. E., Painter, T. H., Dangel, S. & Martonchik, J. V. (2006). Reflectance quantities in optical remote sensing-definitions and case studies, *Remote Sensing of Environment* **103**(1): 27–42.

Scheuer, E. M. & Stoller, D. S. (1962). On the generation of normal random vectors, *Technometrics* **4**: 278–281.

Schott, J. R. (1997). *Remote Sensing: The Image Chain Approach*, Oxford University Press, Oxford.

Schott, J. R., Salvaggio, C. & Volchok, W. J. (1988). Radiometric scene normalization using pseudoinvariant features, *Remote Sensing of Environment* **26**(1): 1–16.

Sen, A. & Srivastava, M. (1990). *Regression Analysis: Theory, Methods and Applications*, Springer–Verlag, New York.

Shaw, G. E. (1983). Sun photometry, *Bulletin of the American Meteorological Society* **64**(1): 4–10.

Shaw, J., Barnsley, M. J., Hobson, P. D. & Quaife, T. (2000). Validating satellite sensor estimates of land-surface biophysical properties using plant-growth models and up-scaling, *in* P. Fisher (ed.), *RSS2000: Adding Value to Remotely Sensed Data*, Proceedings of the 26th Annual Conference of the Remote Sensing Society (CD–ROM), Remote Sensing Society, University of Leicester, UK, pp. On CD–ROM.

Silva, L. F. (1978). Radiation and instrumentation in remote sensing, *in* P. H. Swain & S. M. Davis (eds), *Remote Sensing: The Quantitative Approach*, McGraw–Hill, London, pp. 21–135.

Slater, P. N. (1980). *Remote Sensing: Optics and Optical Systems*, Addison–Wesley, Reading, Mass.

Slater, P. N., Biggar, S. F., Holm, R. G., Jackson, R. D., Mao, Y., Moran, M. S., Palmer, J. M. & Yuan, B. (1987). Reflectance-based and radiance-based methods for the in-flight absolute calibration of multispectral sensors, *Remote Sensing of Environment* **22**(1): 11–37.

Smith, G. M. & Milton, E. J. (1999). The use of the empirical line method to calibrate remotely sensed data to reflectance, *International Journal of Remote Sensing* **20**(13): 2653–2662.

Smyth, G. K. & Verbyla, A. P. (1996). A conditional likelihood approach to residual maximum likelihood estimation in generalized linear models, *Journal of the Royal Statistical Society Series B (Methodological)* **58**(3): 565–572.

Soares, A., Gomez-Hernandez, J. & Froidevaux, R. (eds) (1996). *geoENV I – Geostatistics for Environmental Applications*, Kluwer, London.

Sprent, P. (1969). *Models in Regression and Related Topics*, Methuen and Co. Ltd., London.

Staenz, K. & Itten, K. I. (1982). Reflective properties of asphalt and concrete surfaces, *Proceedings of the International Symposium of the ISPRS WG-VII*, ISPRS, Toulouse, France, pp. 307–316.

Stein, P. (1999). *Interpolation of spatial data : some theory for kriging*, Springer, London.

Sterne, J. A. C. & Davey Smith, G. (2001). Sifting the evidence – what's wrong with significance tests?, *British Medical Journal* **322**: 226–322.

Stewart, J., Engman, E., Feddes, R. & Kerr, Y. (eds) (1996). *Scaling up in Hydrology using Remote Sensing*, John Wiley and Sons, Chichester.

Strahler, A. H. (1994). Vegetation canopy reflectance modelling – recent developments and remote sensing perspectives, *in* G. Guyot (ed.), *Sixth International Symposium on Physical Measurements and Signatures in Remote Sensing*, International Society for Photogrammetry and Remote Sensing, CNES, Val D'Isére, France, pp. 593–600.

Strahler, A. H., Woodcock, C. E. & Smith, J. A. (1986). On the nature of models in remote sensing, *Remote Sensing of Environment* **20**(2): 121–139.

Strutt, T. H. J. W. l. L. R. (1871). On the scattering of light by small particles, *Philosophical Magazine* **41 (Series 4)**(275): 107–274.

Stuart, A., Ord, J. K. & Arnold, S. (1999). *Kendall's Advanced Theory of Statistics*, Vol. 2A Classical Inference and the Linear Model, sixth edn, Arnold, London.

Subrahmanyam, B., Robinson, I. S., Blundell, J. R. & Challenor, P. G. (2001). Indian Ocean Rossby waves observed in TOPEX/POSEIDON altimeter data and in model simulations, *International Journal of Remote Sensing* **22**(1): 141–167.

Swain, P. H. & Davis, S. M. (eds) (1978). *Remote Sensing: The Quantitative Approach*, McGraw–Hill, London.

Tanre, D., Herman, M., Deschamps, P. Y. & Deleffe, A. (1979). Atmospheric modeling for space measurements of ground reflectances, including bidirectional properties, *Applied Optics* **18**(21): 3587–3594.

Taylor, B. N. & Kuyatt, C. E. (1994). Guidelines for evaluating and expressing the uncertainty of NIST measurement results, *Technical Report NIST Technical Note 1297*, Physics Laboratory, National Institute for Standards and Technology (NIST).

Teillet, P., Markham, B. & Irish, R. R. (2006). Landsat cross-calibration based on near simultaneous imaging of common ground targets, *Remote Sensing of Environment* **102**(3-4): 264–270.

Trotter, C. M., Dymond, J. R. & Goulding, C. J. (1997). Estimation of timber volume in a coniferous plantation forest using Landsat TM, *International Journal of Remote Sensing* **18**(10): 2209–2223.

Upton, G. & Cook, I. (2002). *Dictionary of Statistics*, Oxford University Press, Oxford.

van der Meer, F. (1994). Extraction of mineral absorption features from high–spectral resolution data using non–parametric geostatistical techniques, *International Journal of Remote Sensing* **15**(11): 2193–2214.

Van Gardingen, P. R., Foody, G. M. & Curran, P. J. (eds) (1997). *Scaling-up: From Cell to Landscape*, Cambridge University Press, Cambridge.

Ver Hoef, J. M., Cressie, N., Fisher, R. N. & Case, T. J. (2001). Uncertainty and spatial linear models for ecological data, *in* C. T. Hunsaker, M. F. Goodchild, M. A. Friedl

& E. J. Case (eds), *Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications*, Springer, New York, pp. 214–234.

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modelling: The SAIL model, *Remote Sensing of Environment* **16**(2): 125–141.

Vermote, E. F., Tanre, D., Deuze, J. L., Herman, M. & Morcrette, J. J. (1997). Second simulation of the satellite signal in the solar spectrum, 6S: An overview, *IEEE Transactions on Geoscience and Remote Sensing* **35**(3): 675–686.

Verstraete, M. M., Pinty, B. & Myneni, R. B. (1996). Potential and limitations of information extraction on the terrestrical biosphere from satellite remote sensing, *Remote Sensing of Environment* **58**(2): 201–214.

Walker, A. H., Muller, J. P. & Naden, P. S. (1999). High resolution interferometric SAR DEMs for hydrological network derivation, *Remote Sensing of the System Earth – A Challenge for the 21st Century*, Proceedings of the 1999 IEEE International Geoscience and Remote Sensing Symposium, IEEE.

Walthall, C. L., Norman, J. M., Welles, J. M., Campbell, G. & Blad, B. L. (1985). Simple equation to approximate the birectional reflectance from vegetative canopies and bare soil, *Applied Optics* **24**(3): 383–387.

Wang, J. D., Jiao, Z. T., Gao, F., Xie, L., Yan, G. J., Xiang, Y. Q., Liang, S. L. & Li, X. W. (2003). Validation of MODIS albedo product by using field measurements and airborne multi-angular remote sensing observations, *IGARSS 2003: IEEE International Geoscience and Remote Sensing Symposium, Vols I - Vii, Proceedings - Learning from Earth's Shapes and Sizes*, pp. 1888–1890.

Webster, R. (1989). Is regression what you really want?, *Soil Use and Management* **5**(2): 47–53.

Webster, R. & McBratney, A. B. (1989). On the Akaike information criterion for choosing models for variograms of soil properties, *Journal of Soil Science* **40**(3): 493–496.

Webster, R. & Oliver, M. (2001). *Geostatistics for Environmental Scientists*, John Wiley and Sons, Chichester.

Wen, X. H. & Gomez-Hernandez, J. J. (1996). Upscaling hydraulic conductivities in heterogeneous media: An overview, *Journal of Hydrology* **183**(1-2): ix–xxxii.

Wicks, T. E. (2000). *Driving an Ecosystem Simulation Model with Spatial Estimates of LAI*, PhD thesis, Southampton University.

Wikle, C. K. (2003). Hierarchical models in environmental science, *International Statistical Review* **71**(2): 181–199.

Willink, R. & Lira, I. (2005). A united interpretation of different uncertainty intervals, *Measurement* **38**(1): 61–66.

Wilson, A. K. (1997). An integrated data system for airborne remote sensing, *International Journal of Remote Sensing* **18**(9): 1889–1901.

Wilson, M. & Atkinson, P. M. (2005). The use of elevation data in flood inundation models: a comparison of ERS InSAR and combined contour and differential GPS, *International Journal of River Basin Management* **3**(1): 3–20.

Yang, H., Zhang, J., van der Meer, F. & Kroonenberg, S. B. (2000). Imaging spectrometry data correlated to hydrocarbon microseepage, *International Journal of Remote Sensing* **21**(1): 197–202.

Youden, W. J. (1997). *Experimentation and Measurement*, NIST (Special Publication 672), Washington.

Zha, Y., Gao, J., Ni, S. X., Liu, Y. S., Jiang, J. J. & Wei, Y. C. (2003). A spectral reflectance-based approach to quantification of grassland cover from Landsat TM imagery, *Remote Sensing of Environment* **87**(2-3): 371–375.