UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

School of Geography

Per-pixel uncertainty in change detection using airborne remote sensing

Volume 1 of 2

Kyle Mackenzie Brown

Thesis for the degree of Doctor of Philosophy

April 2005

UNIVERSITY OF SOUTHAMPTON

Abstract

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

SCHOOL OF GEOGRAPHY

Doctor of Philosophy

Per-pixel uncertainty in change detection using airborne remote sensing

by Kyle Mackenzie Brown

The coastal region of the UK is under intense anthropogenic and environmental pressure that results in negative environmental impacts. To manage coastal habitats, monitoring programs are required that can provide indications of where environmental change has occurred. Remote sensing has been recognised as one method of providing these monitoring programs, but operational detection of land cover change is currently not achievable.

This study aimed to develop remote sensing methodologies that could be used for operational monitoring of natural and semi-natural coastal habitats by governmental organisations such as Environment Agency or English Nature. The errors associated with post-classification change detection in the coastal zone using remotely sensed data were researched and methods of deriving thematic and geometric per-pixel uncertainties from airborne sensor data were investigated.

Methods of deriving a per-pixel geometric uncertainty model for the Compact Airborne Spectrographic Imager (CASI) were examined. A correlation was found between angular acceleration of the aircraft platform and geometric errors of automatically geocorrected CASI imagery. This relationship was used in combination with a geometric uncertainty model to provide a per-pixel model of instrument geometric uncertainty. The instrument geometric uncertainty model was combined with a model of orthometric error to provide a probabilistic geometric uncertainty model. A misregistration model was derived from the geometric uncertainty model and a significant correlation was found between the predicted and actual misregistration. Thematic uncertainty measures were derived from the output of the multi layer perceptron (MLP) and probabilistic neural network (PNN). A correlation was found between the thematic uncertainty measures derived and pixel thematic error. Heuristics to maximise the accuracy of the thematic uncertainty measures were derived.

The geometric and thematic uncertainty measures were combined in a model of change detection uncertainty. Using synthetic data and data from a sand dune test site the use of uncertainty measures in change detection was found to be significantly more accurate compared to a change detection model that did not include uncertainty.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

All the remotely sensed data were provided by the Environment Agency. The Ainsdale Sand Dunes data was acquired for "The Development of Remote Sensing Techniques for Marine SAC Monitoring", a collaborative project between the Environment Agency and English Nature.

Many thanks to the following:

Giles Foody for his support, supervision and humour during the whole traumatic PhD experience.

Peter Atkinson for his support, leftfield questions and awareness of the spatial dimensions. I apologise for not having a hint of a variogram in this study.

The Environment Agency who funded the project. Everybody at the Environment Agency who supported me, particularly Paul Williams and Dave Palmer. Special thanks go to Chris O'Dwyer, the senior remote sensing surveyor, who provided me with a great deal of additional information and acquired excellent data under difficult conditions. Many thanks to Alison Matthews for proof reading this thesis.

Everybody at English Nature who helped out, particularly Rob Wolstenholme, the warden of the Ainsdale Sand Dune National Nature Reserve who put up with a lot of questions. Thanks to Elly Hill who was in charge of the English Nature side of the "The Development of Remote Sensing Techniques for Marine SAC Monitoring" project.

My family.

Dave O'Neill and all my other friends for their support.

Nigel Marsh for proof reading the thesis. Its 3-3, but I think you win on penalties.

Madeline for being there and putting up with me. Thank you.

ABBREVIATIONS USED

AGL	Above ground level
CASI	Compact Airborne Spectrographic Imager
DEM	Digital elevation model
dGPS	Differential global positioning system
DSM	Digital surface model
EA	Environment Agency
EC	European Community
EML	ERDAS Macro Language
EN	English Nature
GCP	Ground control point
GPS	Global positioning system
HRV	High resolution visible
IDL	Interactive Data language
IMU	Inertial measurement unit
LiDAR	Light Detection and Ranging
ML	Maximum likelihood classifier
MLP	Multi layer perceptron neural network
NDVI	Normalised Difference Vegetation Index
PNN	Probabilistic neural network
PPS	Pulse per second
RMSE	Root mean square error
SAC	Special Area of Conservation
SPOT	System Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
TM	Thematic Mapper

1 Introduction

1.1 Introduction

Within the UK the coastal region is under intense pressure due to requirements for agricultural, housing, commercial and recreational land. Anthropogenic and natural forcings result in a variety of negative environmental impacts in the coastal zone including erosion, pollution, eutrophication, and the drainage and reclamation of intertidal areas. In recent years the environmental importance of conserving coastal habitats has been recognised and certain of these habitats are now protected under the 1992 European Community Habitats Directive (EEC, 1992). The Habitats Directive requires that monitoring of these protected habitats and reporting of their extent and condition is carried out every five years. However, sea level rise and anthropogenic pressures on the coastal zone have created an environment where changes occur at a wide variety of temporal and spatial scales. This range of scales means that ground-based resources available for coastal monitoring may be inadequate to cope with the large areas that require coverage and may not provide all of the information required. Remote sensing has been used for coastal mapping of habitats such as saltmarsh and sand dunes (Gross *et al.*, 1986; Donoghue and Shennan, 1987; Gross *et al.*, 1987; Donoghue *et al.*, 1994; Hobma, 1995; Thomson, 1995; Fontana *et al.*, 1997; Zhang *et al.*, 1997; Brown and Arbogast, 1999; Shanmugan *et al.*, 2003; Thomson *et al.*, 2003), as well as change in those habitats (Smith *et al.*, 1998; Seeliger *et al.*, 2000). These studies indicate that there is the potential to provide an overview of the changes taking place within coastal habitats at national scales using remote sensing.

Over the last twenty years sensor technology has advanced a great deal with the introduction of fine spatial and spectral resolution imaging sensors mounted on both aircraft and satellites. There have been advances in sensor technology, huge increases in processing power and memory capabilities and increases in the accuracy of automated classification and other techniques. Despite these advances, operational methods of monitoring that can determine changes in surface cover type have not been developed and further study is required to develop the use of remote sensing for monitoring land cover change (Donoghue, 2002; Song *et al.*, 2002; Parr *et al.*, 2003).

Understanding and quantifying the errors within remotely sensed data is essential to reduce the impact of these errors on the accuracy of change detection. Recent studies have begun this process, examining the errors and uncertainties within geographical information sciences and

remote sensing and their propagation when datasets are merged or modelling is carried out (Stanislawski *et al.*, 1996; Steele *et al.*, 1998; Pontius, 2000; Carmel *et al.*, 2001; Crosetto and Tarantola, 2001; Atkinson and Foody, 2002; Aires *et al.*, 2004, Comber *et al.*, 2004; Liu *et al.*, 2004; Mahapatra *et al.*, 2004). The work reported in this study built on previous work on error and uncertainty within remote sensing and developed it further, aiming to provide operational methods of detecting land cover change in coastal habitats.

This study focused on methods of coastal land cover change detection that would be suitable for two organisations, the Environment Agency (EA) and English Nature (EN). The EA is the single largest flood defence authority in England and Wales and has a statutory requirement to protect the environment (Environment Agency, 2005). The coastal zone is of particular interest within the EA, as prevention of flooding and environmental protection can sometimes have conflicting needs. The purpose of EN is to promote the conservation of England's wildlife and natural features (English Nature, 2005). EN has a statutory duty to provide indicators of the status of various coastal habitats and is also involved in the management of various protected sites within England (English Nature, 2005). For these two organisations, monitoring is required to provide the information on which coastal management decisions can be made, as well as for statutory reporting.

Within EA and EN the requirement for accurate, precise and repeatable monitoring of land cover type has been recognised and the potential of remote sensing has been identified (Brown *et al.*, 2003a). This study builds on a collaborative project between EA and EN, called "The Development of Remote Sensing Techniques for Marine SAC Monitoring", which identified that further work is required to develop operational change detection methodologies for the coastal zone (Brown *et al.*, 2003a). The study aimed to develop operational methods, suitable for use by EA or EN, of predicting change in the land cover of natural and semi-natural coastal habitats, such as saltmarsh and sand dunes, using remotely sensed data. Of particular interest were methodologies that could determine the extent and category of change in land cover type and identify where these changes were taking place.

Within this study there is a tension between the requirements to develop methodologies that had the greatest potential to result in the most accurate outputs and those that were suitable for reporting for the EA and EN. As the methodologies developed had to be suitable for operational monitoring for EA and EN, methodological compromises were made during the study which may not have been made if the aim of the study was to achieve the most accurate change detection using remote sensing. These compromises are discussed in the relevant sections.

1.2 Change detection

Many remote sensing studies involve the mapping of land cover or land use. Though baseline land cover or land use data are important, it is detecting change in land cover or land use that is ultimately of the greatest managerial, legislative and scientific interest. Some of these methods predict changes in the area of land cover or land use classes. Predictions of the total change in area of different land cover types may be suitable for some legislative reporting, but for management decisions an output is required that provides indications of where change has taken place and so only methods that could map change were examined in this study.

Several methods have been proposed for change detection using remote sensing that provide various indicators of change. However, studies have generally concentrated on three broad techniques (Mas, 1999):

1. Spectral change.
2. Multitemporal image classification.
3. Post-classification analysis.

1.2.1 Spectral change

Spectral change detection generally involves the use of vegetation indices (Nelson, 1983; Townshend and Justice, 1995; Lyon *et al.*, 1998; Yuan *et al.*, 2002; Nordberg and Evertson, 2003) or change vectors (Lambin and Strahler, 1994; Johnson, and Kasischke, 1998; Melgani *et al.*, 2002; Chen *et al.*, 2003a; Liu *et al.*, 2004).

An index such as the normalised difference vegetation index (NDVI) (Tucker, 1979) that can be correlated with a variable of interest such as green biomass may be used for change detection. On a per-pixel basis the indices from two time periods are subtracted so that an NDVI difference image is formed (Lyon *et al.*, 1998; Nordberg and Evertson, 2003). This indicates areas where changes in productivity or phenology have taken place and the magnitude of the changes.

The change vector is the difference within feature space of a given pixel over time. The magnitude of the vector indicates the magnitude of the change and the direction of the vector indicates the type of change that has taken place. This approach may be used to detect gross

changes in land cover type, such as deforestation, desertification, flooding or fire scars (Lambin and Strahler, 1994; Melgani *et al.*, 2002). It is a simple model of change and may be used to provide an indication of the magnitude of change within feature space. However, spectral change will not necessarily be linked to land cover change (Yamamoto *et al.*, 2001).

If the spectral change method is used, a threshold level that indicates change must be determined and this may be difficult to achieve (Smits and Annoni, 2000; Melgani *et al.*, 2002; Chen *et al.*, 2003a). Though this method can provide indications of change, quantifying that change may be difficult (Lu *et al.*, 2004). The spectral change approach is inappropriate if the datasets that are available have different bandsets (Yamamoto *et al.*, 2001) and it is difficult to account for geometric errors. When the images are merged, geometric errors in either image will result in the images not being precisely co-located in the combined dataset. This means that the combined dataset contains the combined geometric errors, potentially resulting in change detection errors (Townshend *et al.*, 1992; Zhan *et al.*, 2002). Accurate radiometric normalisation is also required, otherwise changes will be predicted due to the changes in radiance rather than changes in reflectance. Accurate radiometric normalisation may be difficult to achieve if the images are obtained at different times of year (Mas, 1999). Data need to be acquired at similar times of year as the spectral characteristics of surfaces change due to seasonal changes in vegetation (Coppin *et al.*, 2004; Lunetta *et al.*, 2004). The spectral change method also does not predict land cover classes at t_1 and t_2, only general classes such as deforestation or no change (Lu *et al.*, 2004).

This project aimed to provide a method of detecting changes in land cover type for EA or EN monitoring programs and it is likely that methodologies would be applied for long term monitoring. Therefore, there may be a requirement to use data from different sensors and acquire the data at different times of year. This linked with the need to provide estimates of the amount of cover for all the land cover types present, means that spectral change techniques were inappropriate for this study.

1.2.2 Multitemporal image classification

Multitemporal classification involves combining two images and classifying the combined image to provide a thematic map of land cover change (Hame *et al.*, 1998; Bruzzone and Prieto, 2000; Bruzzone *et al.*, 2004; Guild *et al*, 2004). In remote sensing, classification is the process by which sensor data are converted to thematic maps usually representing land cover or land use

classes. In the case of multitemporal image classification, the thematic maps represent change classes. The classification process is described in Appendix B.

Generally the multitemporal classification approach uses an unsupervised classification to identify areas of change (Hame *et al.*, 1998; Mas, 1999; Bruzzone and Prieto, 2000). There is only one classification carried out using this method and so the potential effect of classification errors will be less than a method that uses two classifications (Lu *et al.*, 2004). As with the spectral change method there may be problems with data from different sensors, especially if the bandsets or spatial resolutions are different (Smits and Annoni, 2000). Geometric errors within each image are difficult to account for, as the combined image incorporates misregistration errors (Serra *et al.*, 2003).

If the output change classes are required using this method then prior knowledge of the change within the scene will be needed (Coppin *et al.*, 2004). In theory, knowledge would be required of all possible combinations of change that have taken place within the scene (Serra *et al.*, 2003). The number of change classes for which ground data are required is potentially the number of classes squared, as every class at t_1 has the potential to become any of the other classes at t_2. Though this approach may be suitable in areas where the dynamics of change are well understood, it is not a robust approach for habitat surveillance, as it is unrealistic to have prior knowledge of where change will take place at the start of a monitoring program.

For a practical operational method of detecting change it is not realistic to expect prior knowledge of where change will occur and what that change will be. The difficulty in obtaining ground data for all land cover changes, combined with the lack of knowledge of where change was taking place, mean that the multitemporal classification approach was unsuitable for operational remote sensing within this study's parameters.

1.2.3 Post-classification analysis

The other major approach for change detection is to classify images of different dates and then compare the classifications (Arzandeh and Wang, 2003; Dewidar, 2004; Mehner *et al.*, 2004; Narumalani *et al.*, 2004). For this technique, no prior knowledge of change is required. As the different classified images are merged after the classification, geometric errors may be accounted for at the merging stage. Classifications from different sensors may be used (Yang and Lo, 2002; Serra *et al.*, 2003; Narumalani *et al.*, 2004), though additional analysis may have to be carried out if the spatial resolutions are different.

There are limitations to this approach, as errors within either classification have the potential to create errors in change detection (Coppin *et al.*, 2004; Lu *et al.*, 2004). If only thematic errors are considered, the maximum theoretical error within the final change layer is the sum of the errors of the two classifications. Even if the errors within the classifications are small, when combined for change detection, the error of the change detection layer may be relatively large. An acceptable classification accuracy limit of 85% has been suggested (Wright and Morrice, 1997), but in some cases it may not be possible to achieve this due to issues such as ground data availability and spectral separability of the classes used.

To provide accurate measures of change using this technique, the errors within the process should be modelled. Dependent on the classification method, it is possible to provide global and per-pixel probabilities of the error within each classification in order to model and potentially reduce change detection errors (Gong *et al.*, 1996; Shi and Ehlers, 1996; Ediriwickrema, and Khorram, 1997; de Bruin and Gorte, 2000; McIver and Friedl, 2001; Liu *et al*, 2004).

1.2.4 Comparison of change detection methods

Of the three main methods of change detection outlined above, the spectral change method is unsuitable for thematic change detection, being more suited for detection of productivity changes. To determine what change has occurred, the multitemporal image classification method requires knowledge of areas where change will take place prior to the change occurring and is therefore unsuitable as an approach for long term monitoring. Post-classification change detection may propagate and potentially magnify errors and uncertainties within either of the input classifications, but it is possible to incorporate and account for thematic and misregistration errors in the final analysis. Therefore, the post-classification method was most suited to this study.

1.3 Defining land cover change

In the process of determining whether change has taken place the question "What is change?" should be answered. Land cover change occurs in two forms, conversion and modification (Jansen and Gregorio, 2002). Conversion occurs when the land cover or land use changes from one class to another, for example from forest to grassland. Modification occurs when the land cover or land use class remains the same, but there is change within the class, for example the

productivity of woodland may be reduced. The purpose of this study was to examine changes in land cover type and so the priority was to identify land cover conversion.

Using post-classification change analysis, land cover conversion may be defined in two ways (Lu *et al.*, 2004):

1. Whether change has occurred. This may be defined using two classes: Change and No change.
2. What type of change has occurred. For a pixel this may be defined by the class at t_1 and the class at t_2.

The methodologies derived in this study were analysed in the context of the two definitions of land cover conversion above.

1.4 Error and uncertainty in change detection

In a post-classification framework the change detection process involves classifying at least two images from different times that have spatial overlap, co-registering the images and overlaying one classification on the other to determine thematic change. For any change method, the accuracy of the final change detection output will be affected by the following factors (Lu *et al.*, 2004):

1. The accuracy of the input data such as classifications including the availability and quality of ground data to train the classifier.
2. The accuracy of the co-registration.
3. The spatial distribution of the classes. The more heterogeneous the area being monitored, the larger the probability that errors will occur (Serra *et al.*, 2003; Lu *et al*, 2004).
4. The change detection algorithms used.

The change detection process is subject to various errors at each stage of the data acquiring, classification and data merging processes and these errors may be modified, magnified or reduced at each stage of the process (Veregin, 1996; Shi *et al.*, 2004). Therefore, it is essential to track changes in errors as data are changed from low level forms such as imagery to high level abstractions such as themes (Gahegan and Ehlers, 2000). If these errors are not quantified

then the accuracy of the final change map will not be known (Gahegan and Ehlers, 2000; Agumya, and Hunter, 2002), a key factor if decisions are to be made on the basis of this map.

This study examined methods of detecting land cover change that could be used to inform management decisions. These decisions should be based on the causes and effects of any land cover changes. To determine these causes and effects there is a requirement to identify what change has occurred where. By modelling change at the local, per-pixel level it is possible to predict where change has occurred as well as how much change has occurred. To minimise the error associated with per-pixel change detection, the error associated with change predictions should be modelled at the per-pixel level. It is effectively impossible to quantify the actual error on a per-pixel basis for a whole scene, therefore, the focus of any per-pixel error study should be on the probability of error or the uncertainty associated with accuracy of a given pixel.

Uncertainty may be associated with factors such as the geometric positioning of the dataset or the class allocated to a pixel and occurs in two forms: ambiguity and vagueness (Atkinson and Foody, 2002).

Ambiguity is the uncertainty associated with crisp sets and is most commonly expressed as a probability that a variable is correct (Atkinson and Foody, 2002). For example thematic uncertainty may be expressed as the probability that a pixel has been allocated the correct class during a hard classification or geometric uncertainty may be expressed as the probability that a pixel has a geometric error less than one metre.

Vagueness relates to the aspects of fuzzy data models (Atkinson and Foody, 2002). In remote sensing, a fuzzy dataset is generally one in which a pixel can simultaneously be a member of more than one class (Zhang and Foody, 2001; Lo and Choi, 2004; Liu and Wu, 2005) or one in which the boundaries between objects have non-zero width (Cheng and Molenaar, 1999; Zhang and Kirby, 1999; Cheng, 2002; Shi and Liu, 2004). However, there is some confusion in the literature, as in papers by Schneider (1999), Atkinson and Foody (2002), Ahlqvist *et al.* (2003), Hagen (2003) and Wei and Zhang (2004) vagueness is variously described as:

1. An indication of fuzziness.
2. The uncertainty associated with fuzzy datasets.
3. Both fuzziness and fuzzy uncertainty.

According to (Ricotta, 2004) fuzziness and uncertainty are separate concepts, as a dataset in which multiple class membership occurs for a single object such as a pixel is not inherently uncertain. In terms of thematic fuzziness, a pixel could be allocated with 50% membership to two classes with complete certainty. In this case, a fuzzy data model may be applied to the pixel, but there is no uncertainty associated with allocation of the classes. This may be unlikely, as it will be difficult to determine precisely 50% membership to each class, but is a possibility.

To avoid confusion in this study the term vagueness was not be used. Fuzziness was used to define when pixels had membership of more than one class. Fuzziness was said to increase as pixels became more mixed or as a single class became less dominant. Uncertainty was described as occurring when a variable could contain error but the magnitude of the error was not known. Uncertainty measures were defined as probabilities of error or the confidence intervals associated with a variable.

Measures of uncertainty may be used to define the probability of error in terms of a global variable, for example the probability that any pixel within an image will have a geometric error of between one and two metres. Alternatively, uncertainty may be defined in local or per-pixel terms. This could be the probability that a specific pixel will have a geometric error of between one and two metres.

To estimate the uncertainty associated with change detection, the uncertainty in each stage of the change detection process must be quantified and the propagation of errors through the change detection process modelled (Goodchild *et al.*, 1992; Heuvelink, 2002). As errors are passed from source to derived data, the errors are modified such that the characteristics of the error and therefore the uncertainty associated with it may be amplified or suppressed (Veregin, 1996; Shi *et al.*, 2004). Errors within the final change data set could be due to errors in the remote sensing or ground data sets, errors in the classifier, misregistration errors or due to lack of spectral separability of classes (Song and Woodcock, 2003; Carmel and Dean, 2004).

Within much of remote sensing the pixel is assumed to be an accurate method of modelling the environment, but this assumption may not be valid (Fisher, 1997). The spectral response of a sensor varies spatially within a pixel (Manslow and Nixon, 2002) and is influenced by surrounding pixels (Townshend *et al.*, 2000). This unequal response generally means that the centre of a pixel influences the spectra received more than the edges. This can result in errors in classifications, particularly if a pixel contains more than one cover class (Manslow and Nixon, 2002) or is surrounded by pixels of a different class (Townshend *et al.*, 2000). In much of

remote sensing, pixels are assumed to contain only a single class (Fisher 1997). However, where mixed pixels occur, the hard model of a pixel may not be valid and therefore errors will exist when this model is used (Arnot *et al*, 2004).

Despite these limitations, the pixel was assumed to be an accurate model of the environment to limit the scope of this study, as problems with using the pixel model are an entire field of study in their own right. Making this assumption on the validity of the pixel model, there are two main types of error associated with the classified data input into the change detection process and therefore two types of uncertainty that have to be modelled (Serra *et al*, 2003; Carmel and Dean, 2004):

1. Thematic: errors in the allocation of a given class to a pixel.
2. Geometric: errors in the position of a given pixel

1.4.1 Thematic errors

Thematic errors are the errors resulting from pixels being allocated to the incorrect class during the classification process. These errors may be due to a variety of factors including sensor errors, atmospheric or bi-directional effects, land cover classes not being represented accurately by the training data or lack of spectral separability of classes (Song and Woodcock, 2003; Carmel and Dean, 2004).

Thematic errors have the potential to result in errors in the prediction of change, but these errors can interact and be modified during change detection. Thematic errors in either classification are likely to result in errors in change detection (Figure 1.1b). However, thematic errors do not necessarily result in change detection errors. For example if the position and class of a thematic error was the same in both classifications then there would not be errors in prediction that change would occur, though the class at t_1 and t_2 would be incorrect (Figure 1.1c).

1.4.2 Geometric error and misregistration

Assuming thematic datasets from two times are used in the change detection process, geometric errors in either dataset are highly likely to result in errors in the final change detection layer. Misregistration error is the positional error resulting from two or more images not being accurately co-located. As geometric error is likely to have a complex, spatially dependent form, misregistration errors are likely to be complex when layers are merged.

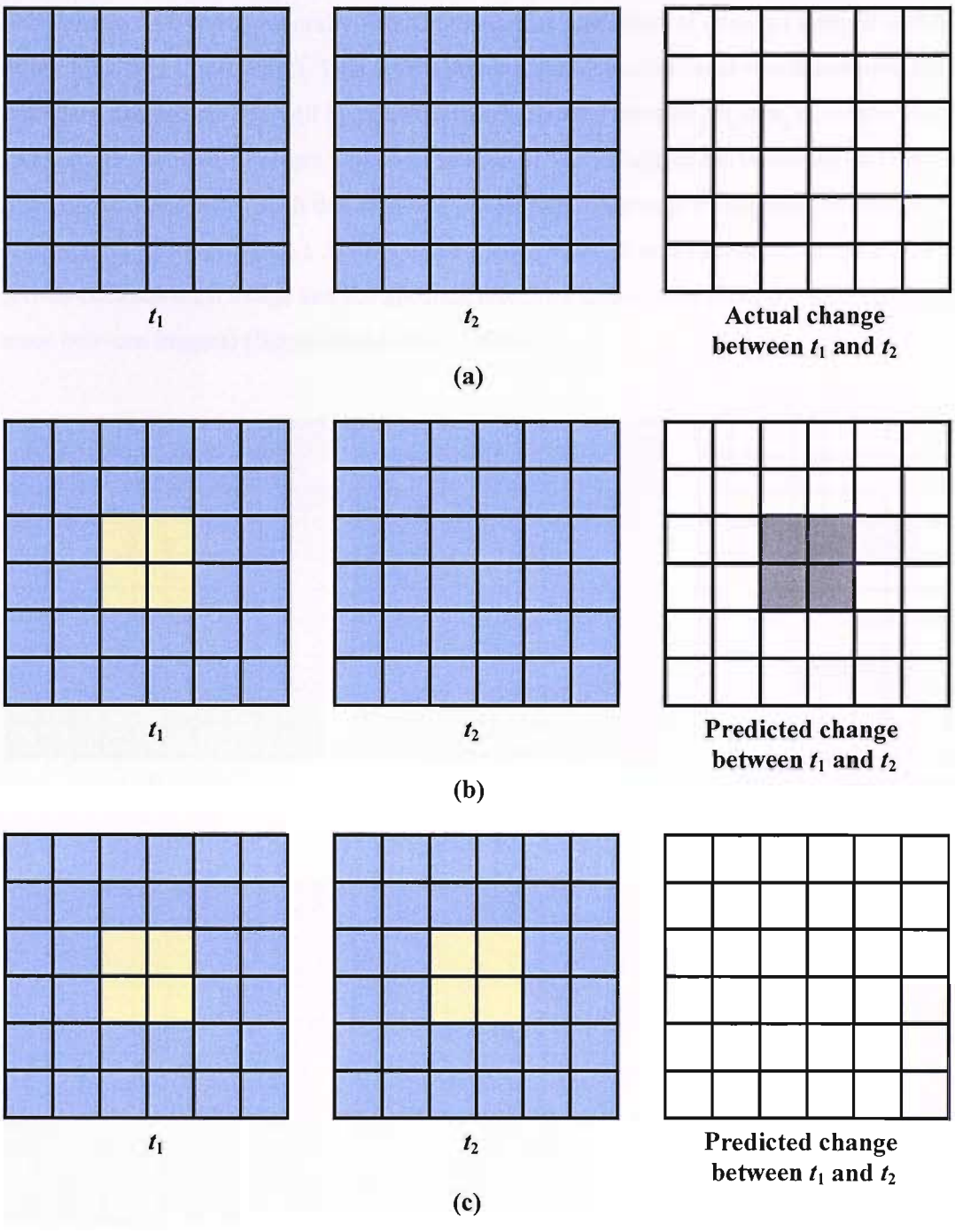


Figure 1.1 Dependency of change detection errors on the interaction of thematic error. Assuming two classes (yellow and blue) and no change between t_1 and t_2, the grey pixels in the change image indicate error due to thematic error.
a) Actual classes at t_1 and t_2.
b) Thematic error at t_1 but not t_2. This results in error in predicted change.
c) The same thematic error at t_1 and t_2 in terms of position and class. This results in no error in predicted change.

If misregistration errors are reduced, the ability to detect change accurately is increased (Dai and Khorram, 1998; Roy, 2000; Chen *et al.*, 2003b). The effect of misregistration errors will be

greatest at the boundary between classes (Dai and Khorram, 1998; Serra *et al.*, 2003; Yamano and Tamura, 2004) and generally results in incorrect prediction of change (Verbyla and Boles, 2000; Stow and Chen, 2002). This means that the spatial resolution at which movement of a boundary can be detected will be related to the horizontal error of the data. However, this is likely to be complex, as errors will be a function of the attitude of the boundary and the x and y component of the error, such that misregistration may incorrectly exaggerate, reduce or obscure change (Roy, 2000) (Figure 1.2). Misregistration errors will be a function of absolute error (errors between each image and the absolute reference frame) as well as the relative error (the error between images) (Stanislawski *et al.*, 1996).

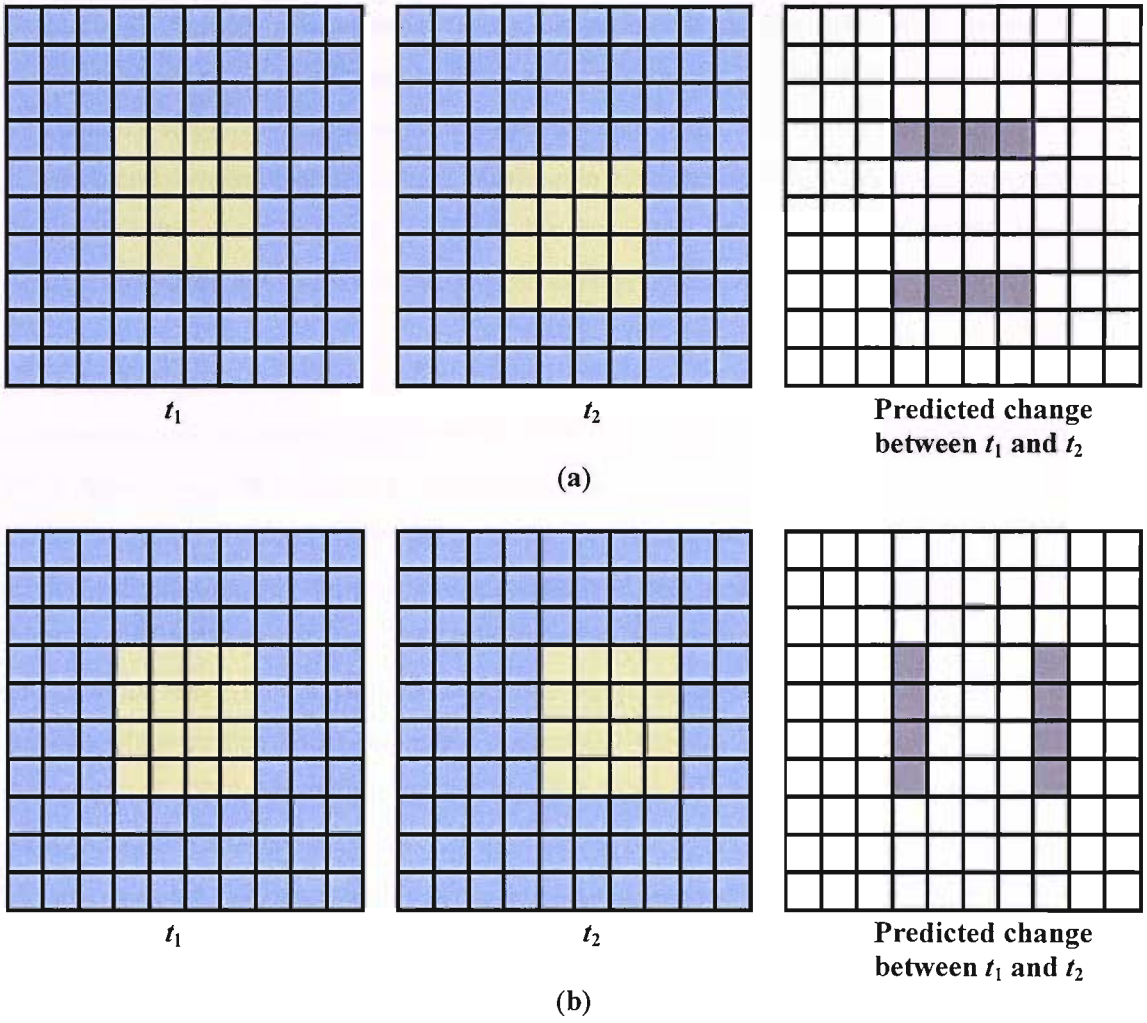


Figure 1.2 Dependency of change detection error on misregistration error direction. Assuming two classes (yellow and blue) and no change between t_1 and t_2, the grey pixels in the change image indicate error due to misregistration. Errors only occur at the boundaries between classes and do not occur if misregistration is parallel to the boundary.
a) Misregistration in north-south direction
b) Misregistration in east-west direction

1.4.3 Uncertainty and error propagation in change detection

Modelling of errors in change detection is essential, as thematic or positional errors within classification layers may be magnified or modified as they are propagated through the change detection model and interact with other variables (Arbia *et al.*, 1998). For example, misregistration errors have a larger impact when land cover classes are fragmented rather than clumped (Verbyla and Boles, 2000; Carmel *et al.*, 2001).

Though uncertainty in change detection has been considered in several studies, these studies tend to concentrate on just per-pixel thematic uncertainty (Shi and Ehlers, 1996; de Bruin, 2000; de Bruin and Gorte, 2000) or global measures of thematic and misregistration error (Carmel and Dean, 2004). As there will be spatial variation in geometric and thematic errors, the use of global measures of uncertainty could restrict change detection. A global error measure, one which uses a single value to describe error for a whole image, will underestimate error in some positions and overestimate it in others, resulting in a reduced accuracy or imprecise measure of change. For this reason methods that determine the spatial distribution of misregistration and thematic errors are most suitable for per-pixel change detection.

As error propagation and transformation occurs each time a conceptual or physical model is constructed using geographical information systems or remote sensing (Lanter and Veregin, 1992; Arbia *et al.*, 1998; Shi *et al.*, 2004) all the input datasets will contain error. Errors in the source data are transferred when operations such as classification or merging data layers are carried out. This may result in the source and derived data having different error characteristics (Veregin, 1996; Warren *et al.*, 2002). It is, therefore, essential to identify errors and uncertainties within all the data used and model the propagation and modification of the errors throughout the change detection process (Lanter and Veregin 1992; de Zeeuw *et al.*, 1999; Crosetto *et al.*, 2001; Heuvelink, 2002; Carmel and Dean, 2004). By modelling the errors, their propagation and effects it may be possible to reduce the errors in land cover change detection. The errors within datasets and their propagation are discussed in the relevant chapters.

1.5 Aims of the study

This study aimed to develop methods that could be used by EA or EN for operational monitoring of natural and semi-natural coastal habitats using remote sensing. The approach taken predicted the errors and uncertainties associated with change detection in the coastal zone using remotely sensed data. These predictions of error and uncertainty were used to reduce the

errors of land cover change detection. Figure 1.3 shows a diagram of the model of change detection error used in this study. The research examined three main areas:

1. Geometric uncertainty and misregistration.
2. Thematic uncertainty.
3. Combining geometric and thematic uncertainty in change detection.

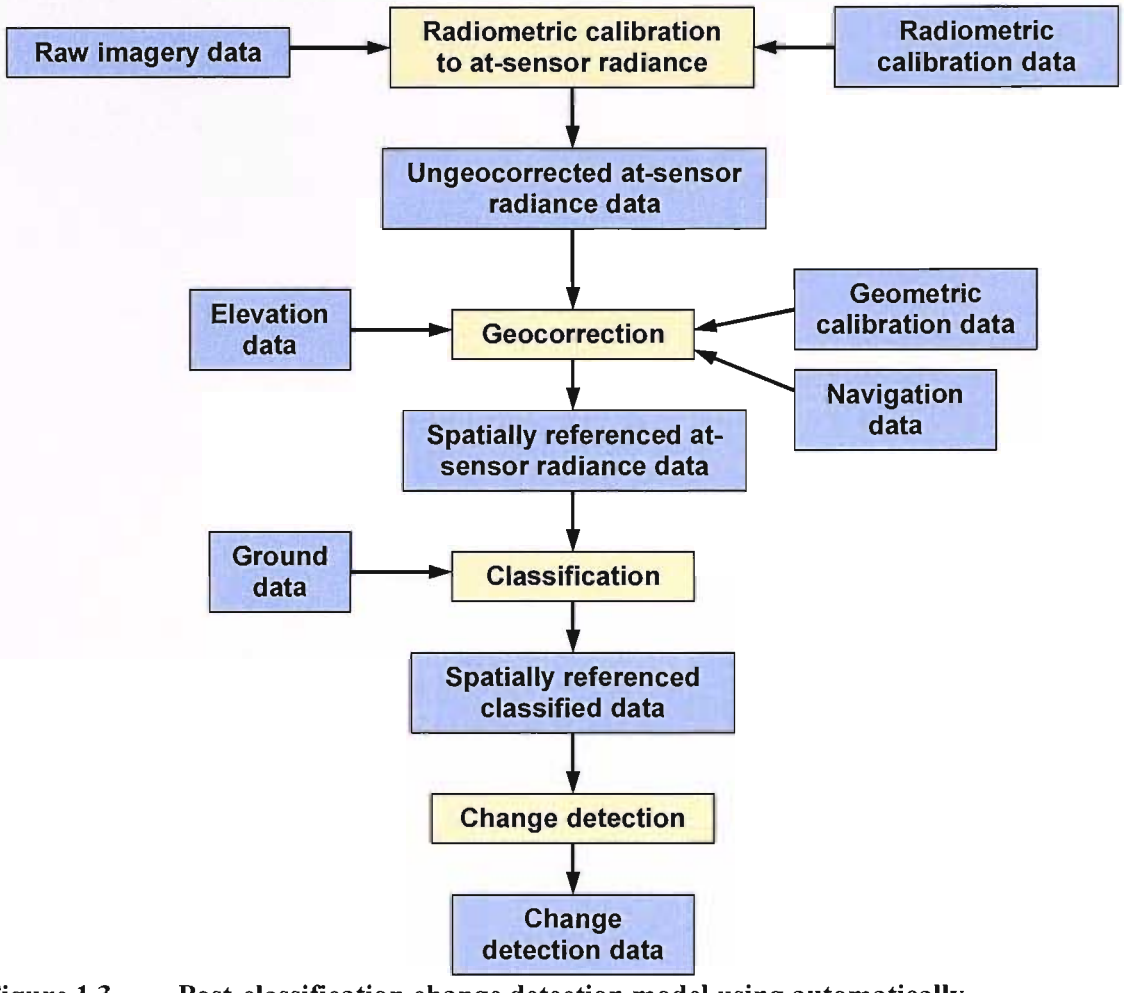


Figure 1.3 Post-classification change detection model using automatically geocorrected imagery. Blue boxes represent data. Yellow boxes represent processes. Data inherently contain errors. Processes can modify those errors.

In Chapter 2 the study sites and remotely sensed and ground data are discussed. Chapter 3 examines the causes of geometric and misregistration errors in airborne data and methods of predicting the magnitude and probability of these errors. To develop a greater understanding the spatial variation in geometric error, the effect of various navigational variables on the geometric accuracy of airborne imagery were tested. These results were used to derive models of the geometric uncertainty and misregistration error. In Chapter 4 methods of deriving thematic uncertainty measures from classifications were examined. The accuracy of these thematic

uncertainty measures was tested and the variables that affected this accuracy were investigated. To determine whether the use of uncertainty could be used to increase the accuracy of change detection, the geometric and thematic uncertainty models were combined in land cover change analysis using synthetic (Chapter 5) and 'real-world' data (Chapter 6). A summary of the research carried out, recommendations for further study and the final conclusions are included in Chapter 7

2 Data

2.1 Introduction

In this chapter the test sites used in this study and the remotely sensed and ground data acquired from the sites are described. All remotely sensed data were acquired by the EA and descriptions are given of the EA's airborne sensors used in this study and the navigational instruments used to geocorrect the sensor data.

A test site at Coventry Airport (Figure 2.1) was used to determine what variables affected the geometric errors associated with airborne imagery. Knowledge of the geometric errors was used to derive geometric uncertainty model that described the per-pixel variation in the probability and magnitude of geometric error. The geometric uncertainty model was used to derive a misregistration model which was tested using data from an urban area of a study site at Ainsdale near Southport, Merseyside (Figure 2.2). An area of sand dunes at Ainsdale (Figure 2.2) was used to test methods of deriving thematic uncertainty models, as well as for assessing the effect of thematic and misregistration uncertainty models on land cover change detection accuracy for a semi-natural habitat.

2.2 Navigational data

To automatically georeference remotely sensed data the position and attitude of the sensor need to be known. In the airborne platform used in this study the position and attitude are provided by GPS (Global Positioning System) and IMU (Inertial Measurement Unit) respectively.

GPS provides x, y, z position data using the Navstar satellite system (Leick, 2004). The time taken for coded signals from several Navstar satellites is measured and used to estimate distance from a satellite. The satellite positions are known and the distance from the satellites may be used to triangulate the position of the GPS receiver. The GPS system used was a combination of the Ashtech Z-surveyor and Novatel MiLLennium®. The specified accuracy of these GPSs when operating on a moving platform is a root mean square error (RMSE) of 10 mm or less plus 1 part per million terms of the separation between the aircraft and the ground GPS systems (Novatel, 1999; ANGEL, 2000).

The IMU provided roll, pitch and heading attitude data. The IMU used was the Applanix POS AV 410 laser ring gyro based on the Litton 200a1 IMU (Brown *et al*, 2003b). The

manufacturer's quoted accuracy is 0.015^{O} RMSE for roll and pitch and 0.08^{O} RMSE for heading (Applanix, 2004).

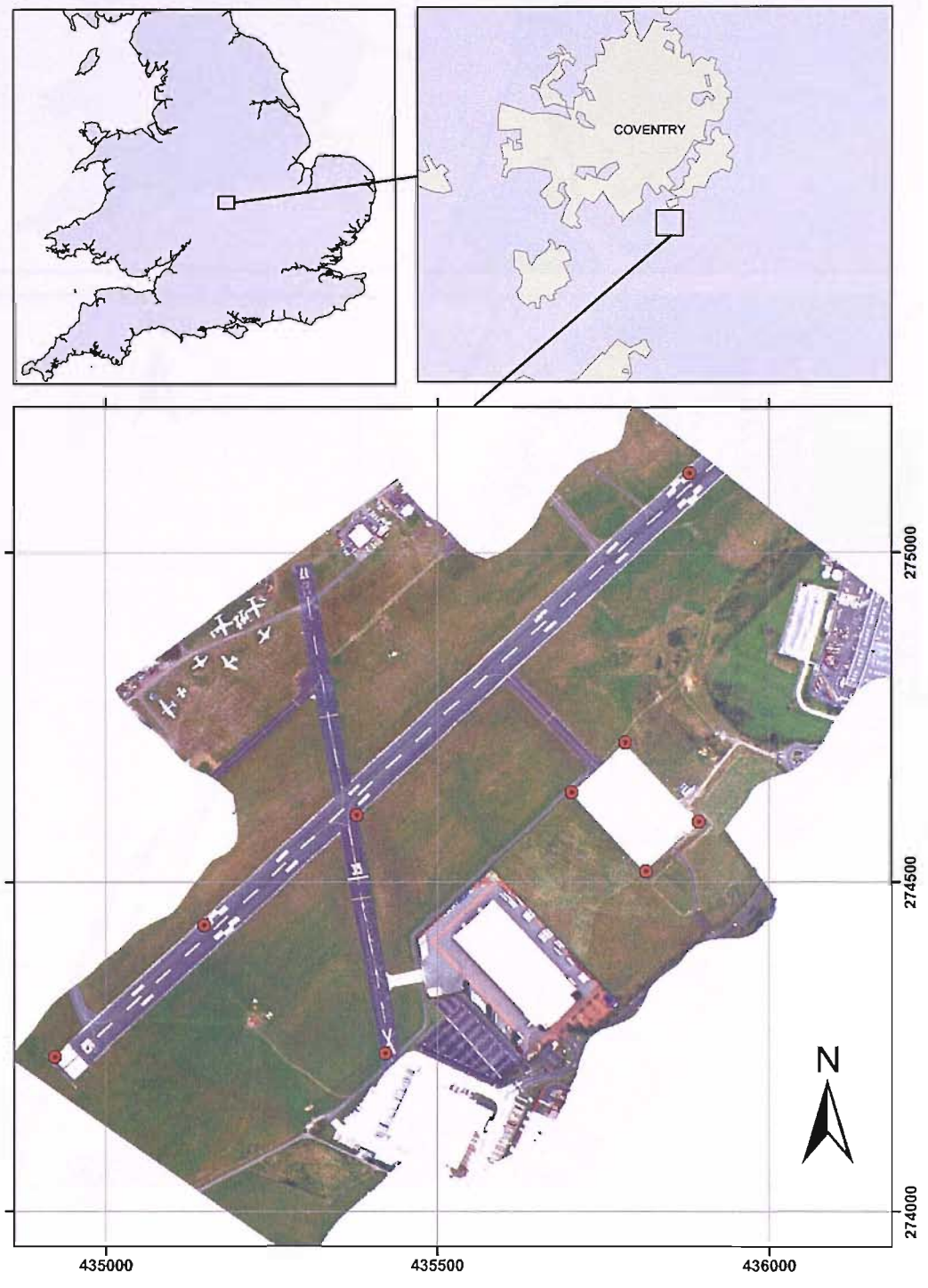


Figure 2.1 Coventry airport test site. Bottom image is true colour CASI imagery of Coventry airport with ground control points overlaid. Grid spacing 500 m.

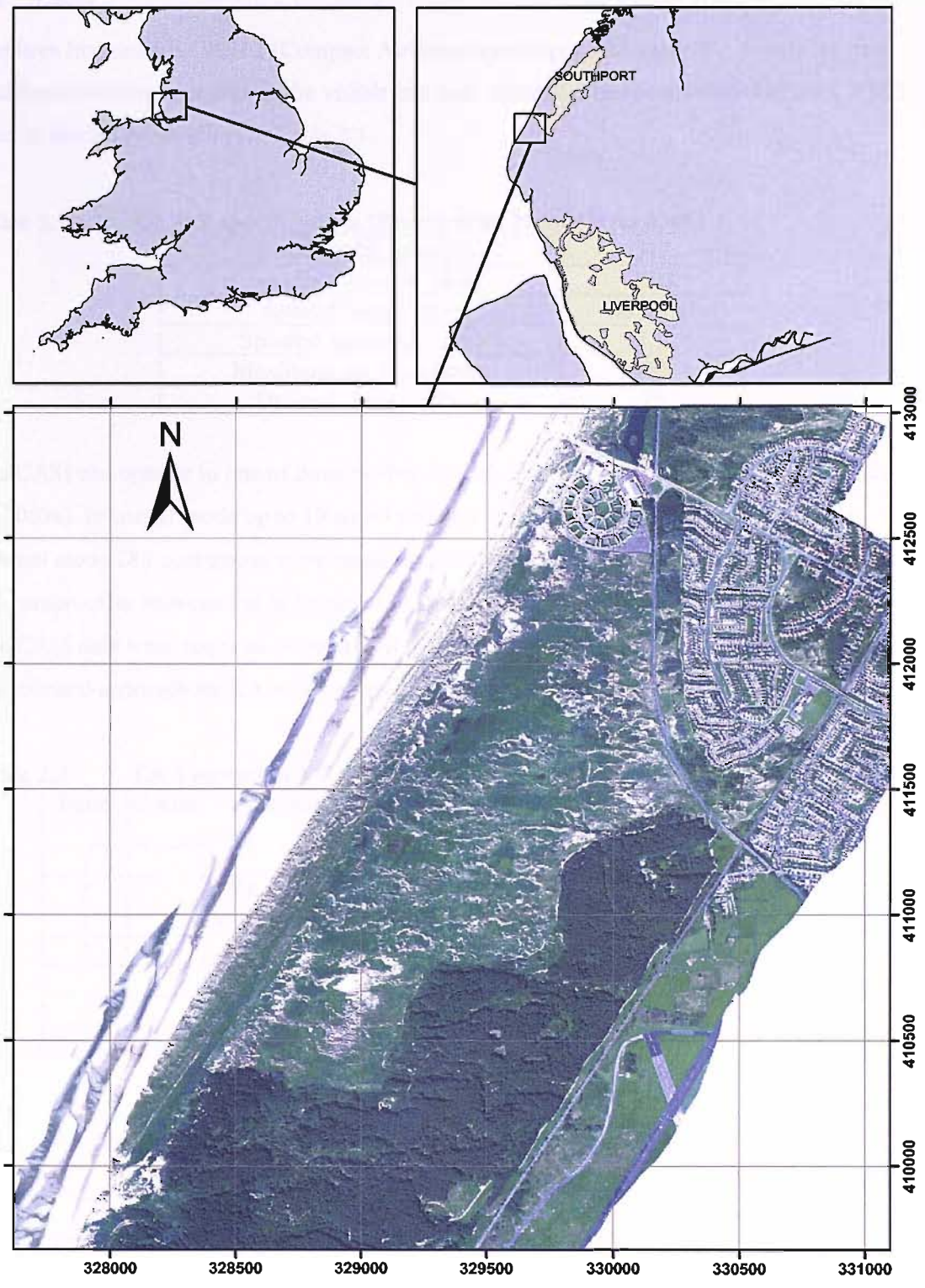


Figure 2.2 Ainsdale test site. Bottom image is true colour CASI imagery of Ainsdale. Grid spacing 500 m.

2.3 CASI

The Itres Instruments CASI 2 (Compact Airborne Spectrographic Imager) is a multispectral pushbroom sensor operating in the visible and near infrared. The specifications of the CASI 2 used in this study are given in Table 2.1.

Table 2.1 CASI 2 specifications (Brown *et al*, 2003b; Itres 2005).

Swath width (pixels)	512
Lens field of view (deg)	37.8
Spectral range (nm)	415-960
Spectral resolution (nm)	1.9
Maximum no. wavebands	288
Dynamic range (bits)	12

The CASI can operate in one of three modes; spatial, spectral and enhanced spectral (Brown *et al*, 2003a). In spatial mode up to 19 wavebands are collected for all 512 spatial pixels. In spectral mode 288 contiguous wavebands are collected for 39 spatial pixels. Enhanced spectral is a compromise between the full number of spatial pixels and the full number of wavebands. The CASI data were acquired using spatial mode and the EA Vegetation 1 bandset (Table 2.2), the standard approach for EA multispectral data acquisition of the terrestrial environment.

Table 2.2 EA Vegetation 1 CASI bandset.

Band	Central wavelength (nm)	+/- (nm)	Band	Central wavelength (nm)	+/- (nm)
1	444.7	5.7	8	710.1	4.9
2	470.8	5.8	9	719.7	4.9
3	490.4	6.7	10	750.4	4.9
4	549.5	5.8	11	761.9	3
5	670.9	5.9	12	780.1	5.9
6	682.4	4	13	860.1	5
7	700.5	4.9	14	880.4	9.8

2.3.1 CASI geocorrection

CASI geocorrection involves the following datasets: CASI imagery, aircraft GPS and IMU attitude data from the aircraft, ground-based GPS position, a digital elevation model (DEM) and geometric calibration data. The calibration data are estimates of the positional offset between the GPS antenna and the CASI sensor and the angular offset between the IMU and the CASI sensor. The navigational data are synchronised with the remotely sensed data using a time stamp sent every second from the GPS, known as a pulse per second (PPS) signal.

The geocorrection process involves the following stages:

1. Post-processing the GPS from the aircraft and a ground-based station. The post-processed data are more precise than the GPS from the aircraft alone.
2. Synchronising the GPS position data, IMU attitude data and CASI imagery using the PPS.
3. Applying the distance offsets between the aircraft GPS and the CASI sensor head.
4. Applying the angular offsets between the IMU attitude and the CASI sensor attitude.
5. Geocorrection of the CASI imagery using the positional and attitude data, a model of the instrument optics and an elevation model.

The GPS post processing is carried out using Waypoint Grafnav (Waypoint, 2001). All other stages use the Itres geocorrection software (Itres, 2000).

2.4 LiDAR

The Optech ALTM (airborne laser terrain mapper) 2033 operates using the principle of Light Detection and Ranging (LiDAR) to derive a digital surface model (DSM). The LiDAR emits laser pulses, which are reflected from a surface and part of the energy returns to the sensor. The time the pulse takes is measured and used to estimate the distance to the surface. The surface the pulse interacts with is generally vegetation, buildings, water or the ground. The distance to the surface is combined with the position of the sensor derived from post-processed GPS and attitude of the sensor derived from the IMU (Section 2.2). The x, y, z position of the pulse on the surface is estimated. The Optech ALTM 2033 operates at a frequency of 33 kHz and uses a rotating mirror to scan across the track of the aircraft resulting in the sawtooth scan pattern in Figure 2.3. All LiDAR data were provided as a 2 m grid resampled using inverse distance weighting.

2.5 Coventry Airport data

To test errors within the CASI geocorrection system, data were acquired over Coventry Airport (Figure 2.1). The Coventry Airport site was used as it is where the EA airplane platform is based, allowing remotely sensed data to be acquired easily and ground control points (GCPs) have been precisely surveyed using GPS, allowing tests to be carried out of the CASI geometric accuracy.

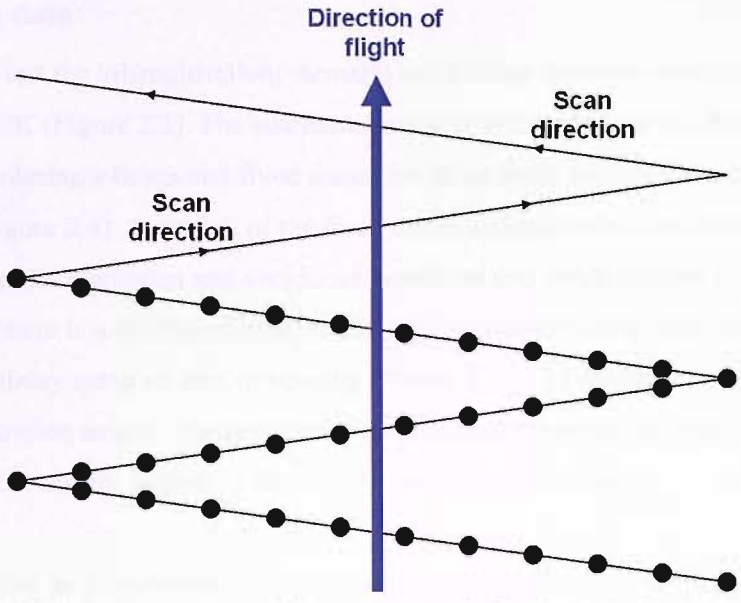


Figure 2.3 LiDAR scan pattern. Dots represent LiDAR pulse footprints. This diagram is not to scale.

13 CASI images with a spatial resolution of 1 m were acquired on 25th August and 9th September 2001 (Appendix C; Table C.1; Table C.3; Appendix D).

Nine GCPs that had been surveyed by the EA were used to test geometric errors of the CASI data (Figure 2.1). The GCP *x, y, z* positions were surveyed using post-processed GPS. The Ashtech Z-surveyor GPS used for the survey has a predicted RMSE error of 5 mm plus 1 part per million for static surveys (ANGEL, 2000). All survey points were within 2 km of the base station and so the RMSE error of the GCPs was estimated to be 7 mm. The points were chosen so that they could be identified easily on 1 m spatial resolution CASI data and consisted of paint markings on tarmac and the corners of artificial surfaces such as tarmac or concrete that were bordered by grass.

To ensure that the elevation used for CASI geocorrection at the GCPs was as accurate as possible, a nearest neighbour interpolation was carried out using the *z*-values of the GCPs and the results resampled to a 10 m grid. By using nearest neighbour resampling the value of the elevation used at the GCPs would be very accurate, minimising orthometric errors, even if horizontal errors occurred. As the GCPs were very precisely surveyed, the orthometric errors at the GCPs, where the horizontal errors were to be tested, would be of a similar order to errors in the GCP positions, which had a predicted RMSE of 7 mm.

2.6 Ainsdale data

The area used to test the misregistration, thematic and change detection models was at Ainsdale, near Southport, UK (Figure 2.2). The southern section of the study area is a dune system with mobile dunes bordering a beach and fixed dunes and dune slack occupy the area inland of the mobile dunes (Figure 2.4). A portion of the fixed dunes and dune slack are wooded, with a mixture of coniferous plantation and deciduous woodland and scrub (Figure 2.4). To the north of the study site there is a developed area, consisting of a holiday camp near the beach and to the east of the holiday camp an area of housing (Figure 2.2). The developed area was used to test the misregistration model. The sand dune site was used to test the accuracy of thematic uncertainty models and the impact of uncertainty on change detection.

The site was chosen as it represented a typical semi-natural coastal habitat that the EA and EN require to be monitored for legislative or management purposes and contained an urban area that could be used to test the misregistration model. The Ainsdale sand dune site was used for "The Development of Remote Sensing Techniques for Marine SAC Monitoring" project mentioned above (Brown *et al.*, 2003a) and is of national and international ecological importance, containing species such as the great crested newt (*Triturus cristatus*), sand lizard (*Lacerta agilis*), natterjack toad (*Bufo calamita*) and the red squirrel (*Sciurus vulgaris*) (WS Atkins, 2004).

2.6.1 Ainsdale remotely sensed data

CASI and LiDAR were acquired over the Ainsdale test site on 28th August 2001 between 1330 and 1400 UT (Appendix C; Table C.2; Appendix D). CASI, LiDAR and true colour digital photography were acquired over the Ainsdale test site on 11th September 2002 between 0940 and 1010 UT (Appendix C; Table C.4; Appendix D). Ideally the 2001 and 2002 data would have been acquired at the same time of day, but this could not be achieved due to weather restrictions. On the day of the 2002 data acquisition a weather window was only available prior to 1030 UT.

The digital photography used in this study was obtained using a Kodak Pro Back ™ Plus 4080 by 4080 pixel CCD mounted on a Hasselblad 205 body with a Zeiss Distagon 40° lens (Brown *et al*, 2003b). The CASI images were geocorrected to a 1 m grid as described in Section 2.3.1 using the LiDAR data as a DEM. The images were mosaiced using ERDAS Imagine and the overlap areas were histogram matched to radiometrically normalise the imagery (ERDAS, 2001). The digital photographs were orthocorrected to the CASI imagery using the LiDAR

DSM and resampled using nearest neighbour interpolation to a 0.2 m grid. To assess the geometric accuracy of the orthocorrection, two independent ground points per digital photograph were derived from the CASI imagery points and were used to derive an RMSE (Appendix F). The orthocorrection geometric RMSE value was 1.80 m.

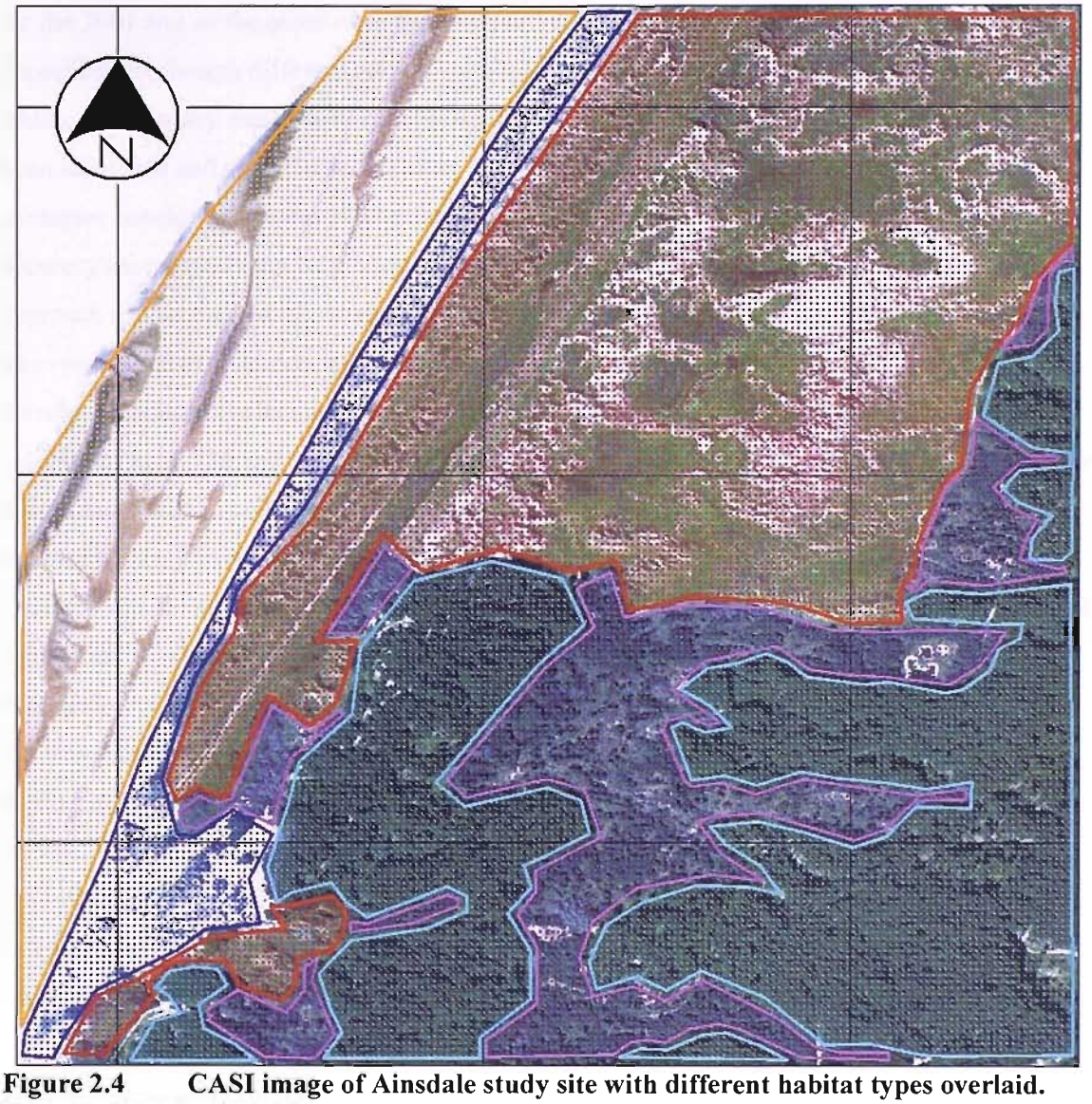


Figure 2.4 CASI image of Ainsdale study site with different habitat types overlaid. Orange = beach; dark blue = mobile dunes; red = fixed dunes; purple = deciduous woodland and scrub; light blue = coniferous woodland. Grid is 500 m.

2.6.2 Ainsdale ground data

Ground data of the Ainsdale site were acquired to train the classifiers, assess the accuracy of the classifications and assess the accuracy of change detection algorithms. The 2001 ground data were collected between 18th and 21st September 2001. The 2002 ground data were collected between 16th and 20th September 2002. Both ground datasets were acquired within 24 days of

the remotely sensed data. It was unlikely that land cover changed between obtaining remotely sensed and ground data.

Contiguous areas of a single land cover class (Table 2.3) were identified during the ground data collection in 2001 and 2002. Additional ground data for accuracy assessment were not available for the 2001 and so the areas were used as training and accuracy assessment data for the 2001 classification, though different areas were used for training and accuracy assessment. In 2002 additional accuracy assessment data points were also acquired in which the dominant class had been identified and so the area data were only used for training the classification. The sampling strategies considered for the 2002 accuracy assessment data are described in Appendix B. The accuracy assessment data were acquired using a geographically stratified random sampling approach as this method allowed a geographical spread of data to be acquired, every pixel had an even chance of selection, a classification was not required prior to sampling and spatial correlation does not affect this method greatly (Appendix B). The strata used were squares with sides 60 m long. The size of the square was selected by estimating the maximum number of data points that could be collected within the study area in the time available for ground data collection. The co-ordinates of the central point of each sample site was obtained using a Garmin GPS II+ differential GPS (dGPS) with a differential signal from the Trinity House Point Lynas station approximately 85 km from the site (Trinity House, 2005). The positions were averaged over a period of one minute. For each sample the ground cover of the classes in Table 2.3 was determined using a 1 m by 1 m quadrat centred on the position measured using the dGPS. Water and sand on the seaward side of the frontal dune system were excluded from the survey. This was due to differences in the tide between the flights and the ground surveying. Therefore, it was not possible to determine the classes at the time of CASI data acquisition using ground based sampling. The class at these points was determined by interpretation of waveband fourteen of the CASI imagery with a central wavelength of 880.4 nm (Table 2.2). Waveband fourteen was used as water has strong absorption at near infrared wavelengths. This should have resulted in more accurate discrimination of sand and water using near infrared than using wavebands in visible wavelengths. The wooded area of the study site also created problems as the differential GPS signal was generally too weak to receive within these areas. In many cases the canopy also restricted the satellite signals to the GPS unit, resulting in a loss of positional accuracy and in many cases the GPS unit could not receive signals from enough satellites to triangulate a position. For these areas, photo interpretation of true colour digital photography was used.

Table 2.3 Thematic classes for ground data collection.

Class name	Class description
Water	Water
Sand	Sand
Marram	Pure Marram grass
Grass	Grasses other than Marram Herbaceous vegetation Moss
Reeds	Reeds/Rushes
Creep	Creeping willow
Buckthorn	Sea buckthorn
Woodland	Deciduous woodland Coniferous woodland

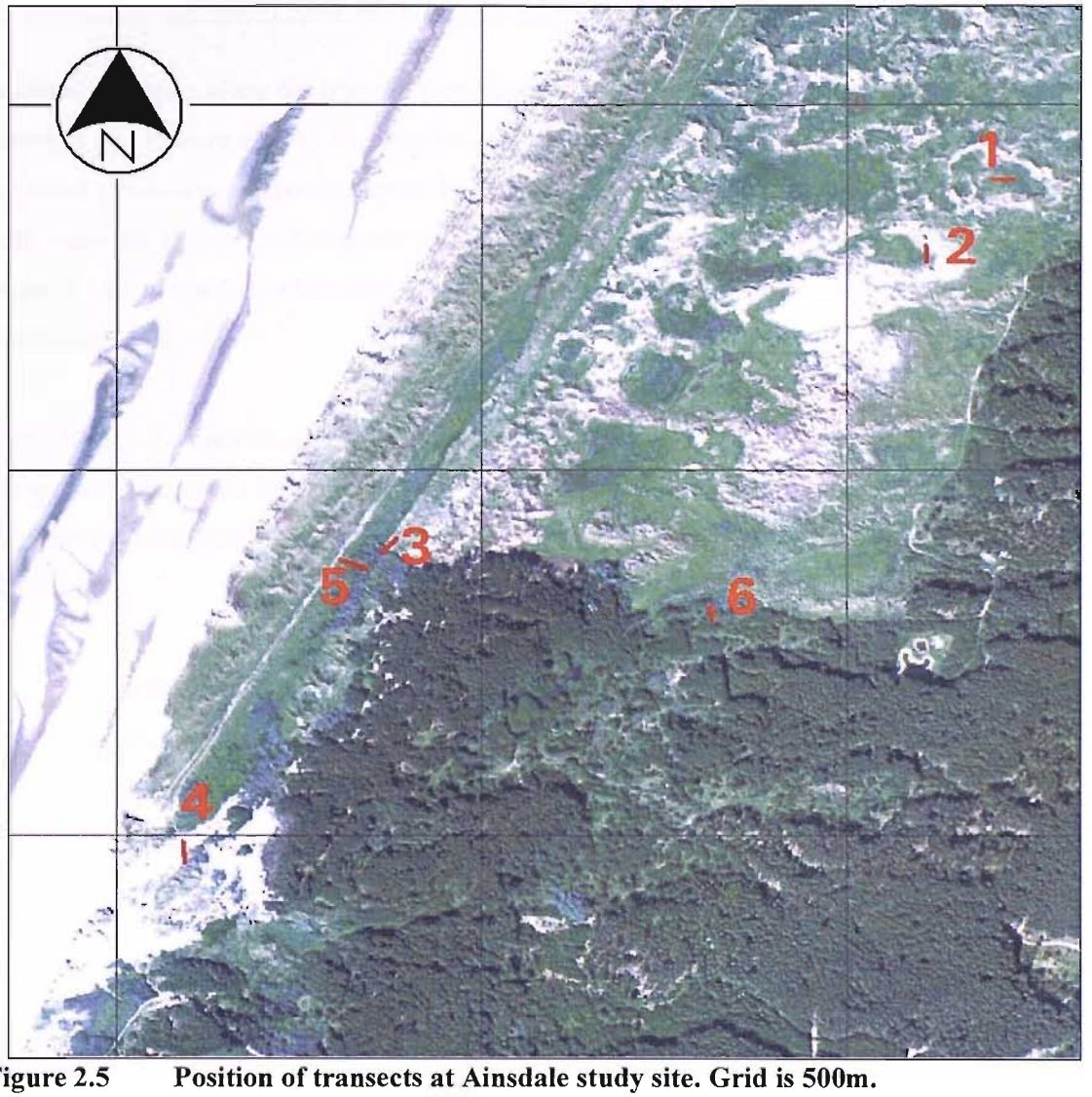


Figure 2.5 Position of transects at Ainsdale study site. Grid is 500m.

Six transect land cover datasets from around the study site were used to test the outputs of the change models. Three were transects taken in areas where there was little or no change, three were in areas where change was very likely to take place between 2001 and 2002 (Table 2.4;

Figure 2.5). Of the three transects where change was likely to occur, one was on the frontal dune system where sand was known to cover vegetation and two were in areas where tree felling was planned (Wolstenholme, personal communication).

Table 2.4 Description of Ainsdale Sand Dunes change transects.

Transect no.	Habitat type	Cover type	Change
1	Fixed dune, dune slack	Grass, Creep	Little/none
2	Fixed dune, dune slack	Grass, Creep	Little/none
3	Fixed dune, woodland	Grass, Woodland	Little/none
4	Mobile dune	Marram, Sand	Sand movement
5	Woodland, fixed dune	Grass, Buckthorn	Buckthorn felled
6	Woodland, fixed dune	Woodland, Grass, Creep	Woodland felled

Positions of points along the transect were surveyed using dGPS. For Transects 2 and 4 the surveyed points were marked by posts prior to the survey. For Transects 1 and 3 one of the surveyed points was marked by a post. For Transects 5 and 6 there were no posts. Additional posts were not allowed and so points were identified for the 2002 survey using dGPS. Along the transects, a 1 m quadrat of the percentage land cover using the eight classes in Table 2.3 was taken every 1 m.

In addition to the transect data for testing the change models, 1 ha square areas of little or no change were identified in conjunction with the English Nature warden during the 2001 and 2002 study period (Wolstenholme, personal communication) (Figure 2.6). Three types of cover were identified:

1. Coniferous woodland (CW): homogeneous stands of coniferous woodland.
2. Deciduous woodland and scrub (DS): heterogeneous areas of deciduous woodland and scrub, with Grass and Creep classes between the trees and scrub
3. Fixed dune (FD): heterogeneous areas of fixed dune and dune slack, with a variety of cover classes including Sand, Marram, Grass, Reed and Creep.

Three sample areas for each cover type were identified. Areas that were excluded from the analysis included the frontal dune ridge in which erosion and accretion were taking place (Brown, 2003a) and the dune area in the centre of the study site. This area had undergone woodland felling between 1992 and 1997 (WS Atkins, 2004) and it was likely that changes due to plant succession were occurring, as well as erosion and accretion of sand. The experiment

assumed that no change had taken place in these areas between 2001 and 2002 and that any change predicted was in error.

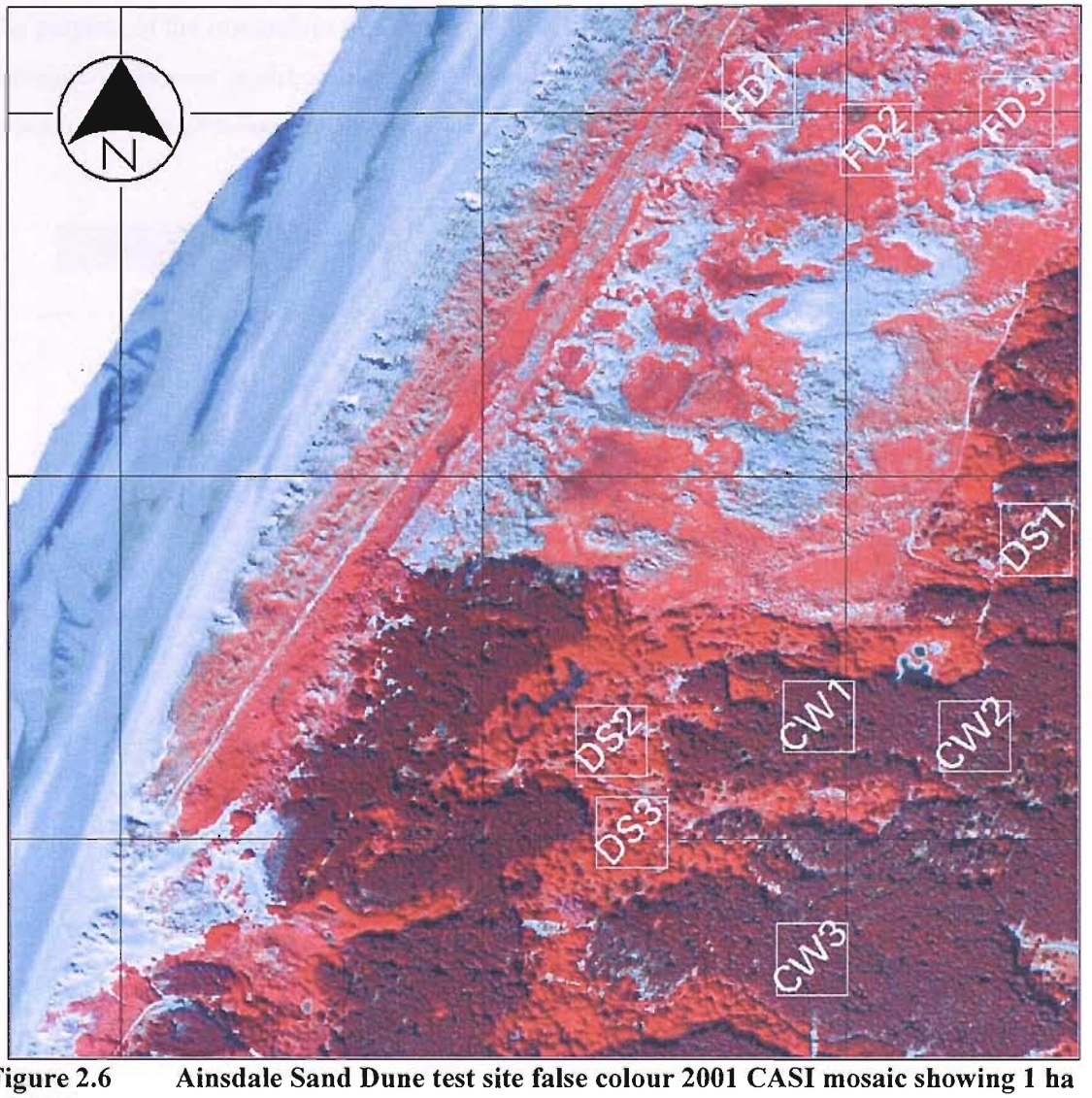


Figure 2.6 Ainsdale Sand Dune test site false colour 2001 CASI mosaic showing 1 ha areas of no change used in testing stage. Grid represents 500 m.

3 Geometric and misregistration uncertainty

3.1 Introduction

The purpose of the research in this chapter was to examine methods of deriving estimates of misregistration error in airborne sensor data and modelling these misregistration errors for use in per-pixel change detection (Figure 3.1).

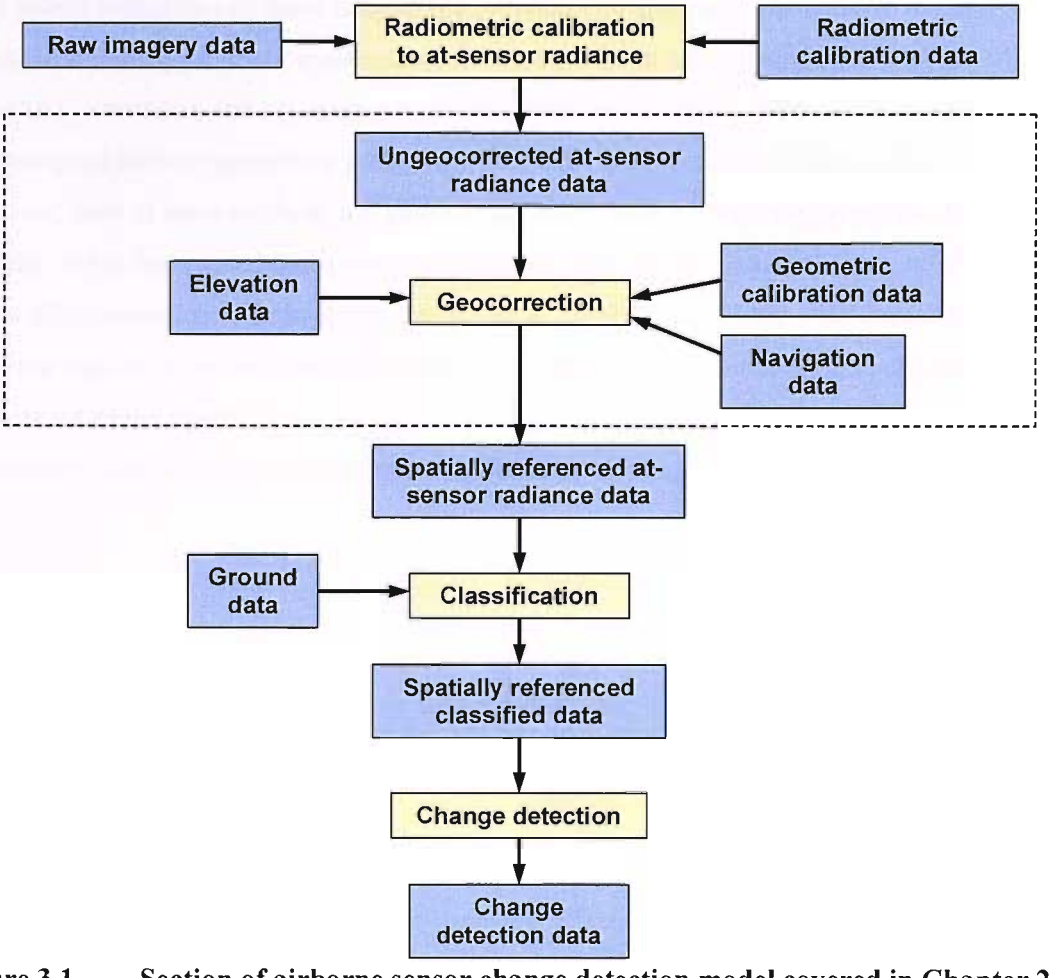


Figure 3.1 Section of airborne sensor change detection model covered in Chapter 2 highlighted in box.

3.2 Estimating misregistration errors

Misregistration errors are a function of the relative error between images and the absolute error of images (Stanislawski *et al.*, 1996) and there are two methods by which these errors may be modelled:

1. By directly estimating misregistration error.

2. By estimating geometric error for individual images and estimating misregistration error from the combined geometric error for two images.

3.2.1 Direct estimation of misregistration error

Misregistration error may be estimated directly by deriving a global measure such as the RMSE (Janssen and van der Wel, 1994). The RMSE is obtained by measuring misregistration error at fixed points within two or more images. If a probabilistic approach was taken to change detection a more appropriate approach would be to generate misregistration probability ellipses from the x, y variance and covariance matrices (Stanislawski *et al.*, 1996). This could be applied on a per-pixel basis to generate a probability of membership based on misregistration errors. However, both of these methods use global error values and misregistration errors often vary spatially. It has been noted that geometric errors are most likely to occur in areas with large terrain differences due to orthometric errors (Stow, 1999). Orthometric errors occur when the elevation used in the geocorrection process is incorrect, resulting in horizontal displacement of the position of the pixel (Figure 3.2). For data acquired using a sensor with a wide swath, orthometric errors are likely to be greatest at the edge of imagery (Figure 3.2).

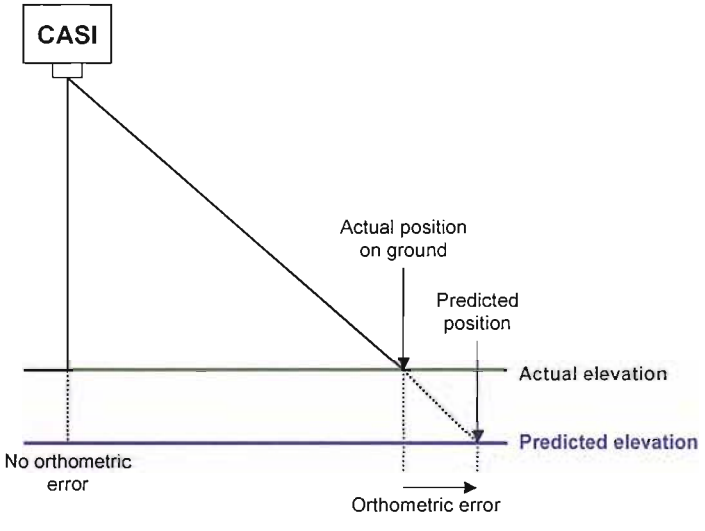


Figure 3.2 Effect of DEM errors on positional accuracy assuming flat terrain.

If a global model of geometric error or misregistration, such as RMSE, was used it would overestimate error in some areas and underestimate it in others. A local measure of misregistration could be used to model misregistration more accurately than a global measure by modelling spatial variation in the error.

One method of estimating the local variation in misregistration would be to estimate the misregistration at a number of points and interpolate between them across the area of interest. This approach would be valid where ground cover does not change or where some areas are known to remain the same. However, there are several limitations with this approach including:

1. In an environment where change is taking place, positional errors may be difficult to estimate if control points are not known to be static.
2. It does not take into consideration local variation in terrain.
3. The edges of images are most likely to exhibit the greatest variation and extremes in misregistration, as the orthometric errors will be greatest.
4. If change detection is carried out using mosaiced images, the geometric errors of the images used to form the mosaic are likely to be different. If this is the case then at the join between images within the mosaic there will be a sudden change in the geometric error vector. Interpolation will not be able to model this error accurately.
5. If misregistration error is derived empirically then values have to be interpolated to obtain a local value. This will smooth predicted misregistration error values and miss extremes of error.
6. There may be difficulty in finding control points from which the misregistration error may be estimated from the imagery. Control points used for this purpose need to be clearly identifiable in all images used and should not move between images. In natural or semi-natural habitats it may be difficult to identify points, as boundaries tend to be indistinct and gradual. This may also be a problem in the coastal zone where the seaward edge of imagery is less likely to have fixed points and often exhibits change due to erosion and accretion.

3.2.2 Indirect estimation of misregistration error

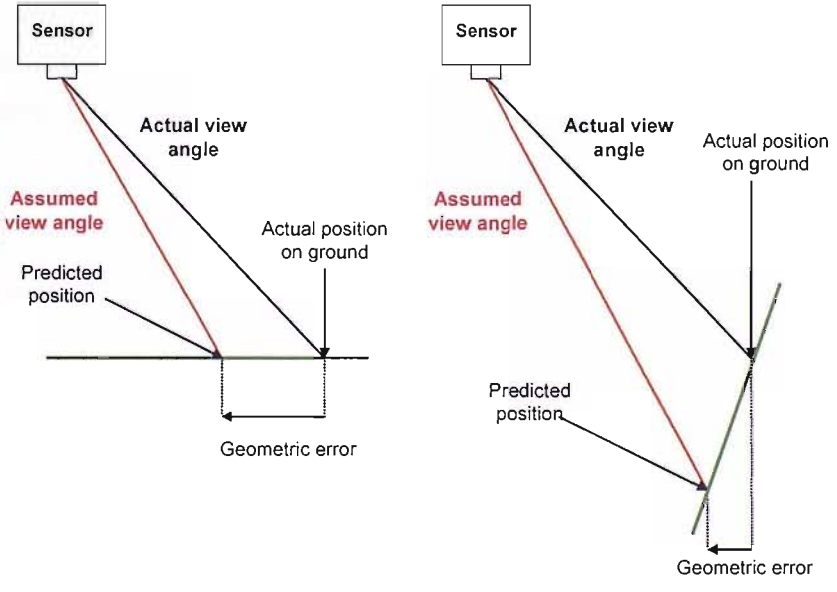
Indirect estimation of misregistration error, by estimating geometric error first, is more complex than the direct approach. However, it can allow estimation of misregistration error to be carried out without using fixed points within two or more images. This is an advantage when semi-natural or natural habitats are being monitored or large amounts of change occur, minimising the number of fixed points for direct estimation of misregistration error.

In automated geocorrection, positional and attitude navigational data are obtained from instruments carried onboard the sensor platform. A geometric model using a terrain model and

the position and the look direction of the sensor, derived from the navigational data, is used to estimate the position of each pixel.

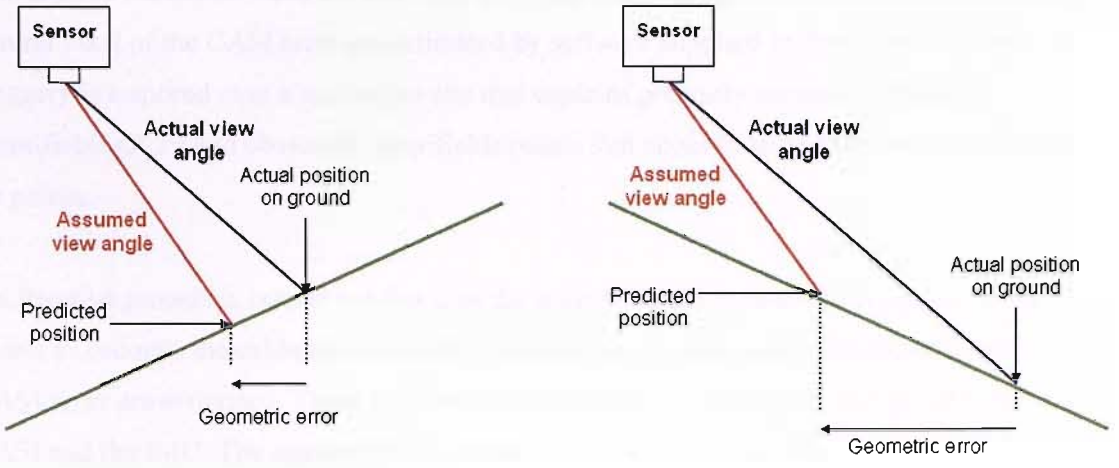
If an automated approach to image geocorrection is used, a geometric uncertainty model could be used to model misregistration uncertainty. Changes in navigational variables such as acceleration have the potential to affect the accuracy of the input data for automated geometric correction (Grejner-Brzezinska *et al.*, 1998; Mostafa *et al.*, 2001; Mostafa and Hutton, 2001). If a relationship between geometric accuracy and variables derived from navigational data were found, it would be possible to derive measures of geometric error without taking measurements based on fixed ground points. This would mean that geometric errors and therefore misregistration errors could be derived in areas where change was taking place or where points were difficult to identify on the ground.

In automated geocorrection, per-pixel geometric errors are caused by errors in the sensor x, y, z positional data, or attitude data or the surface elevation model used. Geometric errors are likely to increase as topographic variation increases (Stow, 1999). As may be seen in Figure 3.3, variations in slope can alter the effect of errors in navigational data on geometric errors. The aspect of the ground can also alter the magnitude and direction of geometric errors (Figure 3.4). If a local measure of geometric error is derived from an automated system without the use of ground control, then terrain effects have to be accounted for.



(a) (b)

Figure 3.3 Example of variation in geometric error due to slope.
a) Slope = 0^{O}.
b) Slope = 70^{O}.



(a) (b)

Figure 3.4 Example of variation in geometric error due to variation of aspect relative to viewing angle. The slope in the two diagrams is the same. Geometric error is largest when aspect is away from sensor.
a) Aspect towards sensor.
b) Aspect away from sensor.

3.3 Misregistration

The magnitude of, and spatial variation in, misregistration errors are dependent on the system used for co-registering images. The Ainsdale study site is coastal, with few obvious fixed points that could be used for ground control and with change occurring particularly near to and below the high water mark (Figure 2.2) (Wolstenholme, personal communication). For these reasons an interpolation approach was not suitable. However, an approach that used an empirical model of misregistration error derived from models of instrument error and orthometric error would be suitable, as ground control in the study site would not be required.

3.4 Calibration of the CASI geocorrection system

Calibration of the CASI geocorrection system has two functions:

1. To calculate the positional and angular offsets between the navigational data and the CASI instrument.
2. To estimate lens focal length and the central pixel of the CASI array.

The *x, y, z* distances between the CASI sensor head and the GPS antenna are measured physically relative to the airframe.

The angular offsets between the IMU and the CASI instrument, the lens focal length and the central pixel of the CASI array are estimated by software supplied by Itres (Itres, 2000). Imagery is acquired over a calibration site that contains precisely surveyed, obviously identifiable GCPs and obviously identifiable points that occur in more than one image known as tie points.

An iterative process is carried out that uses the position of the ground control points and tie points to estimate the calibration variables. Initially, the focal length and the central pixel of the CASI array are estimated. These values are used to estimate the angular offsets between the CASI and the IMU. The angular offsets are then used to recalculate the focal length and central pixel and the process is repeated until each of the variables converges.

3.5 CASI geocorrection system errors

During the geocorrection stage there are the following possibilities for error:

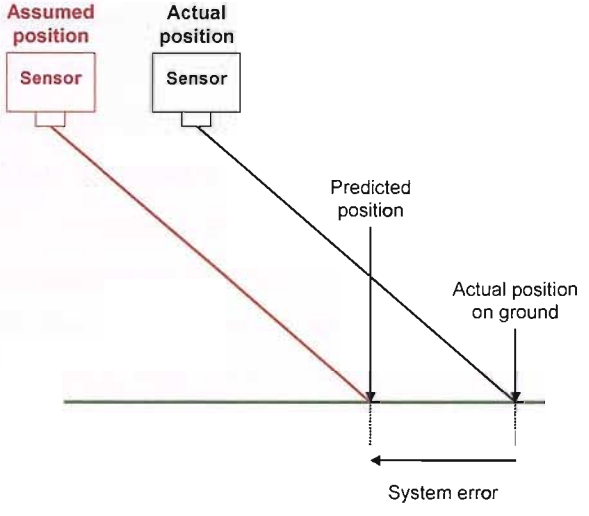
1. Errors in the synchronisation between CASI and the navigational data.
2. Positional errors in GPS.
3. Angular errors in the IMU.
4. Calibration errors in the distance offset between CASI and GPS.
5. Calibration errors in the angular offset between the CASI and IMU.
6. Calibration errors in the focal length of the lens.
7. Calibration errors in the central pixel of the array.
8. DEM errors.

These errors may be split into two groups, instrument errors and orthocorrection errors. Instrument errors consist of points one to seven above, orthocorrection errors will be a function of the instrument errors and the DEM.

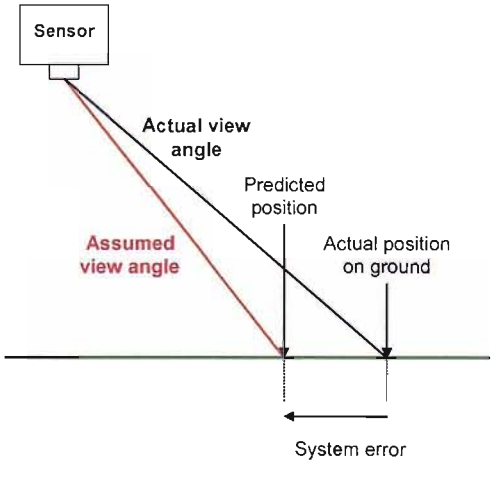
Instrument errors can either be angular errors or positional errors (Figure 3.5). Instrument positional errors result in equivalent positional errors on the ground. Angular errors result in positional errors that depend on the height of the instrument above ground and will be a function of the angular error and height above ground.

3.5.1 Synchronisation errors

Errors in the synchronisation between the CASI and navigational data result in either positional errors or attitude errors. If the synchronisation is offset then positional errors will tend to be along track, as velocity generally changes less than 15% during a flight line (O'Dwyer, personal communication). Attitude errors due to synchronisation errors would be determined by the angular velocity of the IMU. If the angular velocity was small then the attitude used would be closer to the actual value. If the angular velocity was large then attitude would be less accurate. If the synchronisation error is variable then positional errors would be variable, as would attitude errors, but they would also be partially dependent on angular velocity.



(a)



(b)

Figure 3.5 CASI system geometric errors.
a) Horizontal positional system error.
b) Angular system error.

3.5.2 GPS positional errors

Positional errors in the post-processed GPS are dependent on several variables including, the geometry of the satellites, the number of satellites and the quality of the signal received from satellites (Leick, 2004). The GPS system used for positional data in the EA's aircraft (Section 2.2) is quoted as having an RMSE of 10 mm or less plus 1 part per million in terms of the separation between the aircraft and the ground GPS system (Novatel, 1999; ANGEL, 2000). For a typical 20 km baseline this translates to a 30 mm RMSE error. GPS errors result in positional errors.

3.5.3 IMU attitude errors

An IMU consists of accelerometers and gyros (Mostafa *et al.*, 2001). The accelerometers estimate the force applied to the IMU and from this calculate acceleration. The gyro outputs the attitude of the IMU relative to gravitational potential and true north. By combining the post-processed GPS and IMU data it is possible to reduce positional and attitude errors (Hong *et al.*, 2002; Mostafa *et al.*, 2001). However, at the time when the data for this study were acquired, it was not possible to integrate these data streams for all data sets, as key initialisation stages were not carried out when CASI data were acquired. When GPS and attitude data are merged to increase the navigational accuracy, there is a requirement that at least 5 minutes of level flying is carried out, with minimal changes in attitude and velocity. This was carried out when LiDAR data were acquired, but not for 'CASI only' surveys such as those carried out over Coventry airport for calibration and testing purposes. This meant that it was not possible to obtain an integrated IMU and GPS post-processed solution and so this study examines the use of raw IMU data in geocorrection.

Though there are several studies that examine IMU errors and provide a global measure of accuracy (Grejner-Brzezinska *et al*, 1998; Hong *et al.*, 2002; Mostafa *et al.*, 2001), there does not appear to be literature to suggest why errors occur or to link them to environmental effects on the IMU such as acceleration. The specifications of various IMUs provide a maximum acceleration value (Cloud Cap Technology, 2005; BEI, 2005; Novatel, 2005), indicating that this variable will have an influence on accuracy and could be used to provide a local indication of errors within the IMU. Personal experience of the author having geocorrected more than 500 CASI images has shown that positional errors in geocorrected imagery appear to increase during extreme movements of the airframe, particularly roll, pitch or heading changes. From these observations it may be seen that there are two forms of velocity and acceleration that could influence IMU accuracy; linear and angular. Linear velocity and acceleration are in the standard

Euclidean *x*, *y*, *z* coordinates. Angular velocity and acceleration are the velocity and acceleration in units of angle about a point, which in the case of this study was the IMU. IMU errors result in angular errors.

3.5.4 Distance offset errors between CASI and GPS

The distance offset between the CASI lens and GPS antenna is measured directly. According to the EA surveyor responsible for calibrating the system offset between the CASI and the GPS, the maximum measurement errors are less than 15 mm (O'Dwyer, personal communication). Distance offset errors result in positional errors.

3.5.5 CASI system calibration errors

The central pixel of the CASI, the focal length and the roll, pitch and heading offsets are estimated during the calibration process. The central pixel and focal length of the CASI are provided by the manufacturer, but the values are estimates and the errors can be reduced by calibration (Itres, 2000).

Calibration is carried out using four overlapping images acquired at 90° to one another. The image position and *x, y, z* position of the surveyed GCPs are entered into an Itres calibration program (Itres, 2000). In addition, the image positions of easily identified tie points that occur on at least two images are entered into the calibration program. Between 8 and 12 points per image are entered into the program. The calibration program calculates the central pixel of the CASI, the focal length and the roll, pitch and heading offsets.

Any systematic errors in the calibration process should be less than half a pixel (Itres, 2000). Therefore, the combined errors of estimates of angular offset errors between CASI and IMU, the focal length and central pixel should be less than half a pixel. Calibration errors would result in angular errors.

3.5.6 All the CASI system instrument errors

The CASI system errors occur in two forms:

1. Systematic errors: due to calibration errors

2. Non-systematic errors: due to imprecision in the variables used in the geocorrection. These will be errors in the IMU or GPS data.

Positional errors are likely to be a relatively small component of the total, as GPS errors and distance offset errors will be small compared to the 1 m spatial resolution CASI data used in this study. Therefore, it is likely that angular errors make up the largest component of the CASI system geometric errors.

3.5.7 Predicting CASI system errors

A global measure of CASI system errors may be derived by testing for systematic errors, as well as the probability of errors of different magnitudes. To derive a local model of CASI system errors, the errors need to be modelled with variables that can be derived from the system. Easy to derive variables that have a potential to impact on accuracy are linear velocity, linear acceleration, angular velocity and angular acceleration.

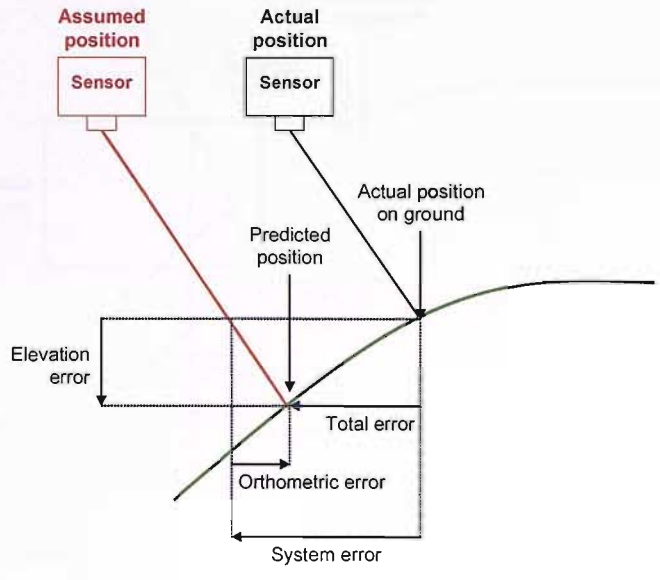
3.5.8 Orthometric errors

Orthometric errors are due to an incorrect elevation being used during the automated geocorrection process. This can be as a result of errors in the DSM or DEM used for geocorrection (Figure 3.2). For this study an Optech model LiDAR was used to derive DSMs (Optech, 2002). The Optech LiDAR system has a quoted error of less than 0.25 m RMSE (Ashkenazi *et al.*, 1999). Errors in the DSM can result in orthometric errors. In automated georeferencing these errors occur when the actual height of the surface being remotely sensed is different from the DEM used and the predicted position of a pixel on the ground is displaced from its actual position. The direction and magnitude of the error is a function of the error in the predicted elevation at a pixel and the look angle of the CASI for that pixel (Figure 3.2).

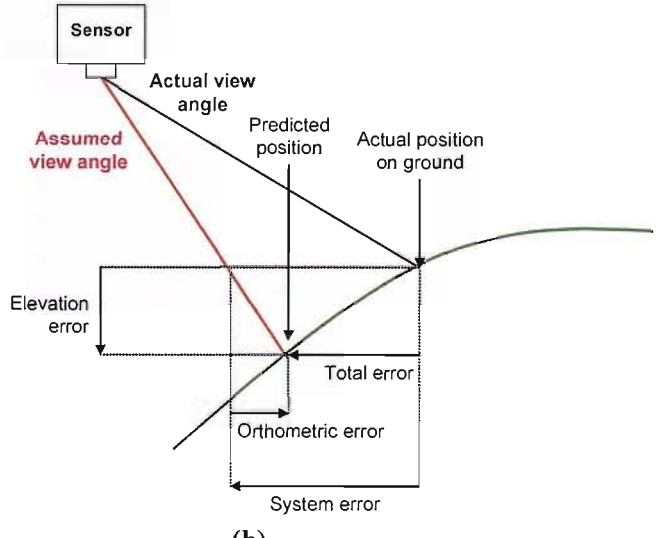
3.5.9 Combined CASI system errors and orthometric errors

The vector of horizontal geometric error is a function of the CASI system error and the orthometric error. However, the CASI system error can influence the orthometric error. Assuming a varying terrain, if the position or view angle of the CASI is incorrect then the elevation extracted from the DEM will be incorrect (Figure 3.6). This will result in an orthometric error, which will be combined with the original instrument error to produce an overall error. The overall error will be a combination of the orthometric and CASI system error

vectors. The magnitude of the total positional error vector may be greater or less than either the orthometric or instrument errors as may be seen in Figure 3.7.



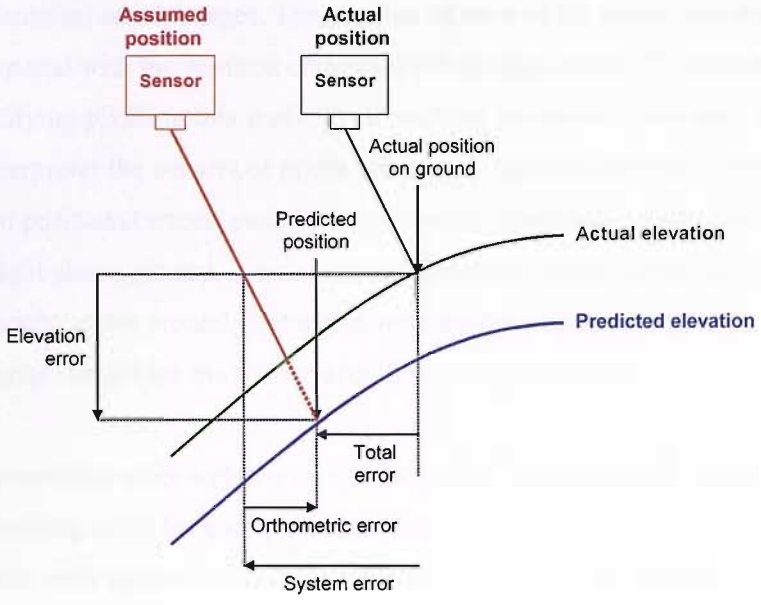
(a)



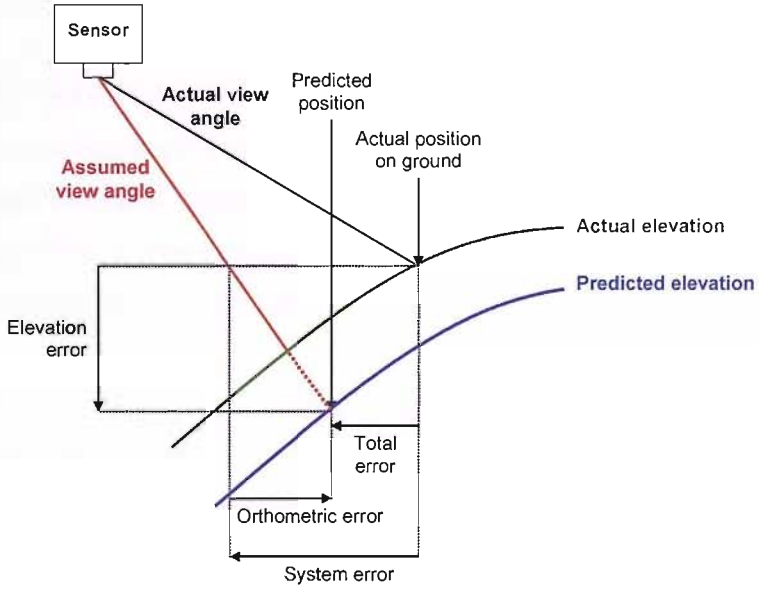
(b)

Figure 3.6 Effect of CASI system errors on positional accuracy assuming varying terrain height.
a) Horizontal positional system error.
b) Angular system error.

The combination of geometric errors due to CASI instrument errors and DEM errors adds an additional layer of complexity to the error model (Figure 3.7).



(a)



(b)

Figure 3.7 Effect of CASI system and DEM errors on positional accuracy assuming varying terrain height.
a) Horizontal positional system error.
b) Angular system error.

3.6 Method and results

To test errors within the CASI geocorrection system, CASI data of Coventry Airport were acquired (Section 2.5). Nine GCPs (Section 2.5) were used to test geometric errors (Figure 2.1).

63 GCPs were identified on 13 images. The position of each of the points was derived from the imagery and compared with the position obtained from ground survey. To have a consistent approach to identifying pixels in this study, in all analysis where positions were derived from imagery by an interpreter the corners of pixels were used. As angular errors were identified as being greater than positional errors, each of the geometric errors was recalculated to account for differences in height above ground. It was assumed that errors were angular and therefore proportional to height above ground. The errors were normalised to a height above ground level of 900 m, the average height for the sorties over the Ainsdale test site.

To compare the geometric error with the various variables being tested the variable of interest was sorted in ascending order for each point and binned in groups of nine to generate the averaged geometric error and averaged test variables such as velocity and angular acceleration. As there were 63 points, bins of nine resulted in an equal number of points per-bin.

3.6.1 Global CASI geocorrection errors

Initial geometric error results gave an RMSE of 1.33 m, a minimum of 0.37 m and a maximum of 2.58 m and a distribution as in Figure 3.8. However, the RMSE and the distribution represent global measures of CASI system geometric error and uncertainty and do not provide an indication of the spatial distribution of geometric errors.

Reductions in CASI geocorrection accuracy due to systematic errors could be compensated for if they could be estimated. Systematic errors could occur in terms of absolute position or relative to flight path. The mean error and standard deviation were calculated in terms of x, y errors and along and across track errors.

The estimated total systematic error in the absolute position was 0.22 m (Figure 3.9). The estimated systematic error relative to the flight path was larger, at 0.43 m (Figure 3.10). This was as expected, as errors in calibration result in systematic errors in position relative to flight direction, rather than absolute position.

Systematic errors were small relative to the size of the pixel, around 30% of a pixel in the along and across track directions (Figure 3.10). Systematic errors were also small relative to the variation in errors, with the across and along track being less than one third of one standard deviation of the error. For these reasons and because the systematic error could not be estimated

for the 2002 data, as CASI data were not acquired over a surveyed test site, systematic errors were ignored for this study.

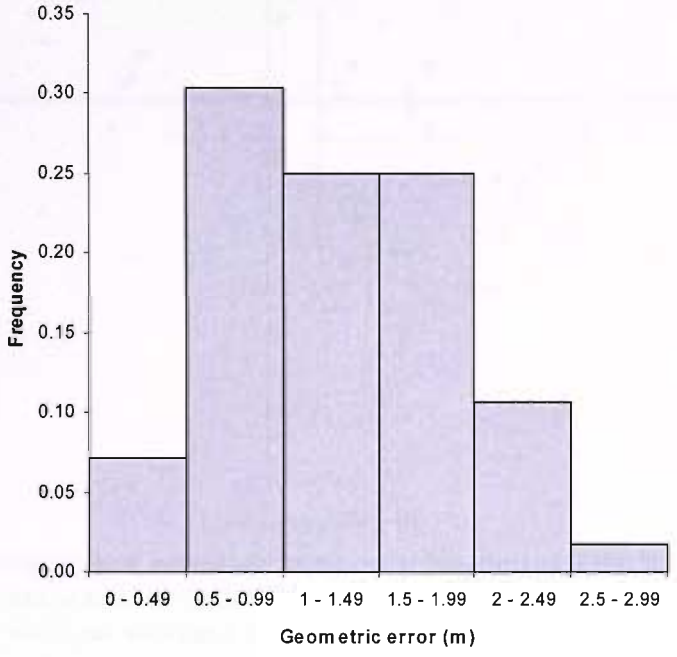


Figure 3.8 Histogram of geometric error frequencies.

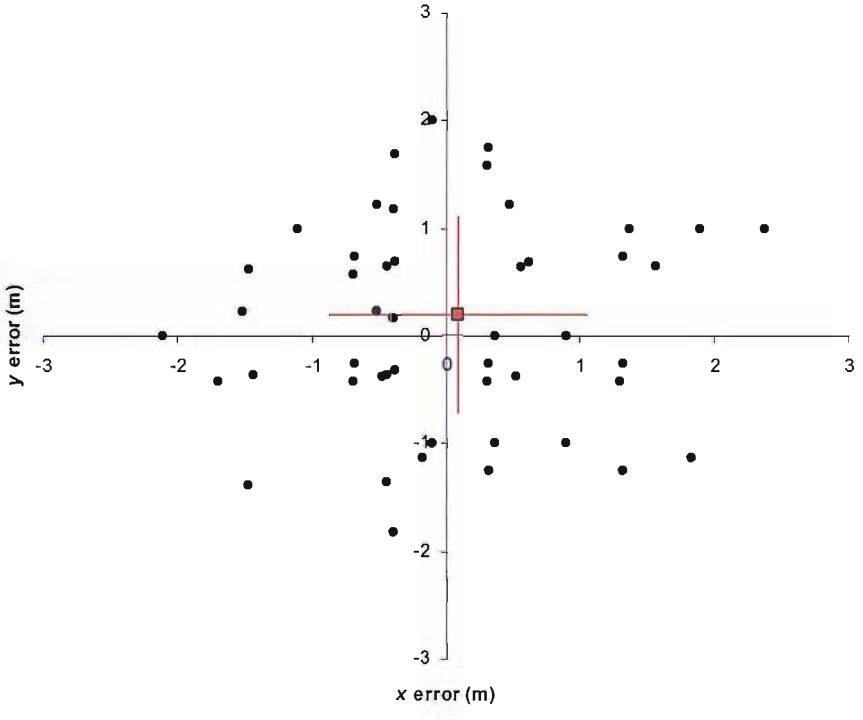


Figure 3.9 Distribution of geometric errors as a function of *x* and *y*. Red box is mean, red lines are one standard deviation. Average *x* error = 0.09 m; average *y* error = 0.20 m; average total error = 0.22 m.

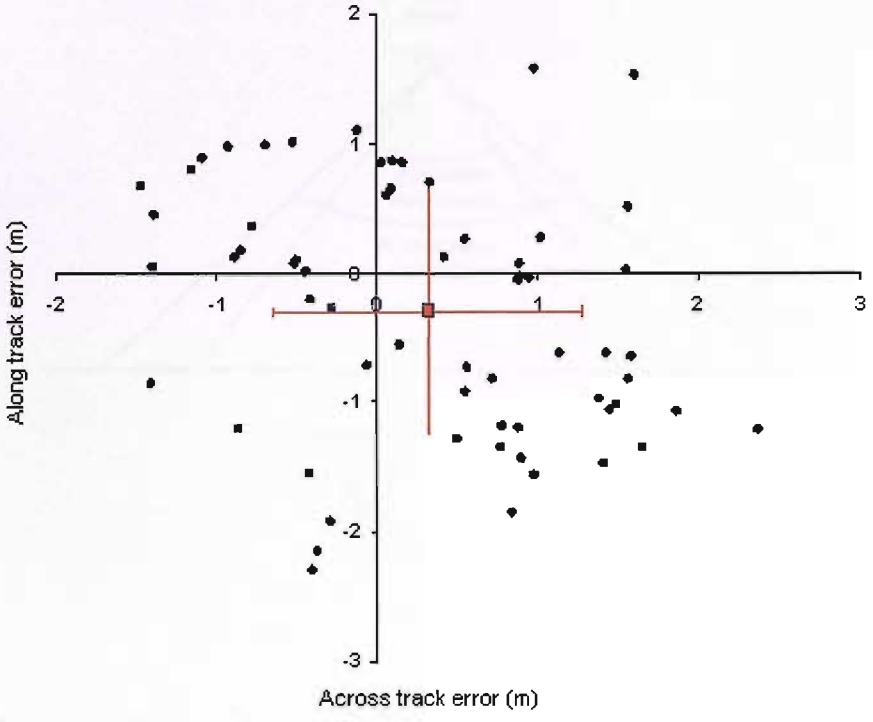


Figure 3.10 Distribution of geometric errors as a function of flight direction. Red box is mean, red lines are one standard deviation. Average across track error = 0.32 m; average along track error = -0.29 m; average total error = 0.43 m.

The geometric error distribution (Figure 3.8) provided a global measure of CASI geocorrection uncertainty, but did not indicate the temporal or spatial distribution of these errors. To determine which variables affected the spatial distribution of geometric errors, linear regression models were fitted to determine whether there were significant correlations between geometric error and a variety of variables.

3.6.2 Lens focal length errors

Errors in lens focal length, resulting from errors during calibration could affect the spatial distribution of geometric errors. Errors in the lens focal length are most likely to occur at the edges of imagery. An incorrect focal length will result in the lens angle being incorrect and this would result in larger errors at the edge of the swath (Figure 3.11).

Errors in the focal length of the lens result in greater geometric errors at the edge of imagery and, therefore, at greater angles off-nadir. To determine whether errors in the estimated focal length of the lens were impacting on geometric accuracy, a linear regression was fitted to determine whether there was a correlation between off-nadir angle and geometric error. The regression showed that the variables were not significantly correlated (Figure 3.12; F=1.39, p=0.292).

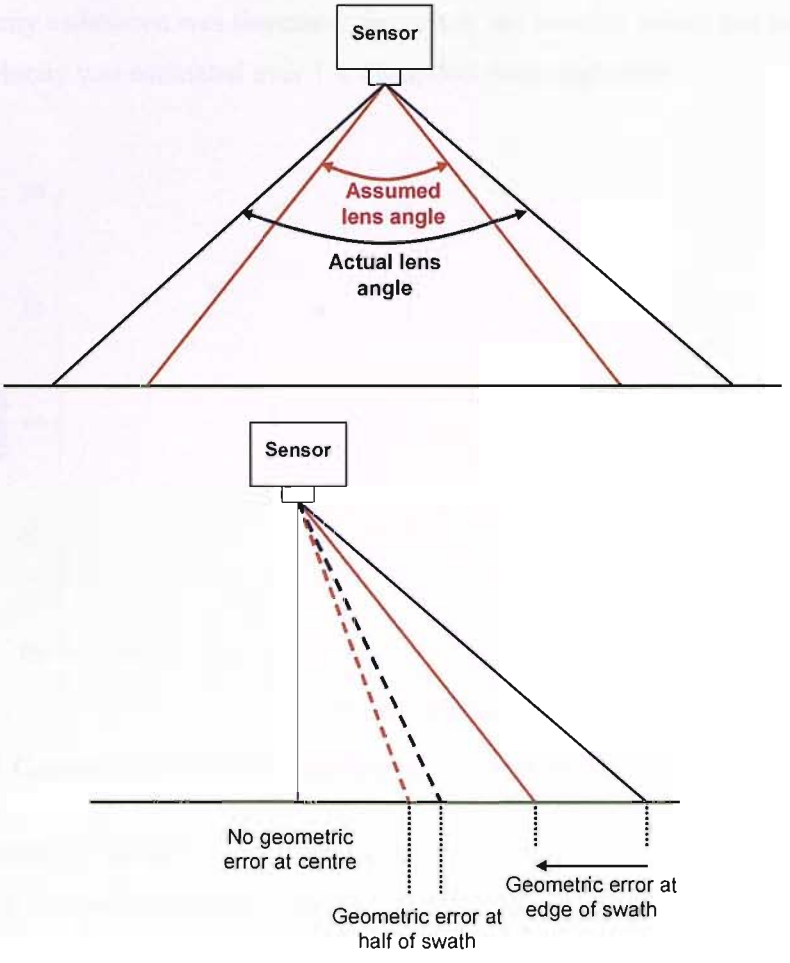


Figure 3.11 Effect of error in predicted lens focal length and therefore lens angle. Errors are greatest at the edge of the swath and least at the centre.

3.6.3 Linking navigational variables with geometric error

The data for estimating navigational variables such as velocity were derived during the Itres geocorrection process, which provided the position and attitude data for every CASI image line. The integration time for all CASI data in this study and therefore the basic time differential between navigational values used was 18 ms.

Velocity (v) was calculated using:

$$v = \frac{\sqrt{(x_{t1} - x_{t2})^2 + (y_{t1} - y_{t2})^2 + (z_{t1} - z_{t2})^2}}{(t_2 - t_1)} \qquad 3.1$$

When the velocity was estimated from the positional data from consecutive image lines, over an 18 ms period, velocity spikes occurred (Figure 3.13), probably due to GPS errors. To reduce these effects, velocity was estimated over a variety of time steps (Figure 3.13). As the time

period for velocity estimation was increased, the size of the velocity spikes was reduced (Figure 3.13). When velocity was estimated over 1 s, the spikes were negligible.

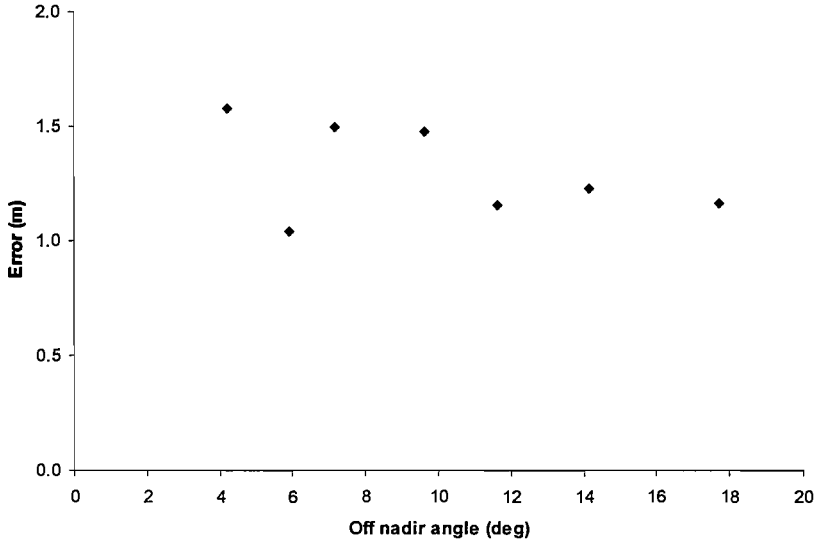


Figure 3.12 Geometric error as a function of off nadir angle.

A linear regression was carried out to determine whether there was a significant correlation between velocity and geometric error. There was no correlation between these variables (at 95% confidence) with all regressions resulting in an R^2 of 0 (Table 3.1).

Table 3.1 Linear regression of velocity against average geometric error. Degrees of freedom were 5.

No steps velocity estimated over	Time linear velocity estimated over (s)	R^2	F	Significance (p)
5	0.090	0.000	0.48	0.519
13	0.234	0.000	0.21	0.666
27	0.486	0.000	0.30	0.607
55	0.99	0.000	0.43	0.541
111	1.998	0.000	0.51	0.507

Acceleration (a) was calculated using:

$$a = \frac{v_{t1/t2} - v_{t2/t3}}{((t_1 + t_2)/2)^2} \qquad 3.2$$

Where $v_{t1\,t2}$ is the average velocity between t_1 to t_2.

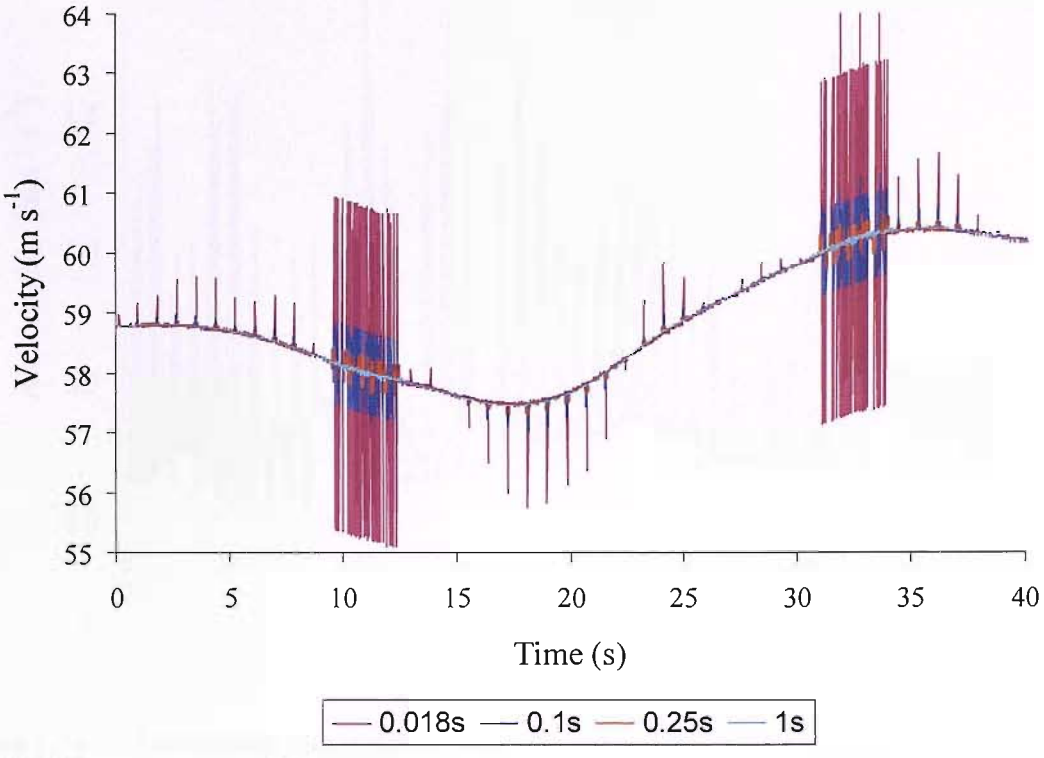


Figure 3.13 Effect of estimating velocity over different time steps. It may be seen that when velocity is calculated from positional data 0.018 s apart, velocity spikes occur. As the averaging time is increased towards 1 s, the velocity spikes become negligible.

This calculation assumes that t_1 - t_2 and t_2 - t_3 are equal. By resolving this equation in x, y, and z it is possible to obtain the following equation:

$$a = \frac{\sqrt{((x_{t1} - x_{t2}) - (x_{t2} - x_{t3}))^2 + ((y_{t1} - y_{t2}) - (y_{t2} - y_{t3}))^2 + ((z_{t1} - z_{t2}) - (z_{t2} - z_{t3}))^2}}{(t_2 - t_1)^2} \quad 3.3$$

As would be expected from the velocity results, noise occurred in the acceleration values when derived over short time periods. By estimating the velocity for the acceleration calculations over longer time periods, the noise was reduced (Figure 3.14).

A linear regression was carried out to determine whether there was a significant correlation between acceleration and geometric error. As may be seen from Table 3.2 the variables were not significantly correlated (at 95% confidence) and had an R^2 of less than 0.013 for all regressions.

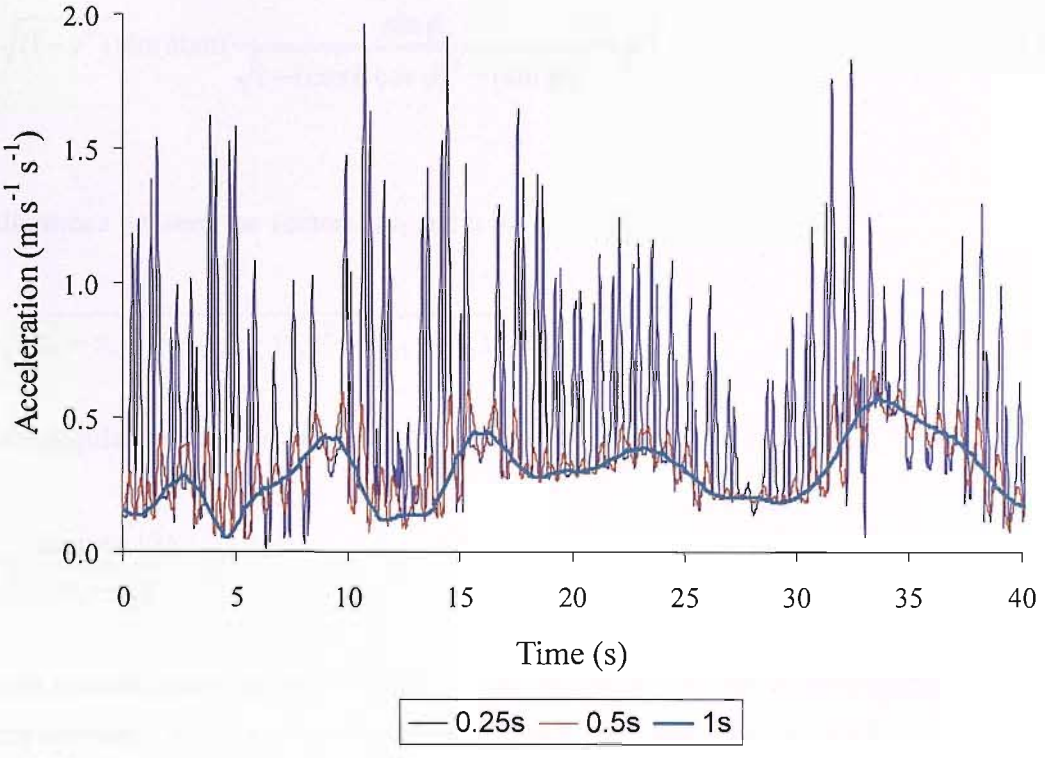


Figure 3.14 Estimating acceleration using velocity calculations over different time steps. It may be seen that when acceleration is calculated from velocity data 0.25 s apart, acceleration spikes occur. As the averaging time is increased towards 1 s, the spikes become negligible.

Table 3.2 Linear regression of acceleration against average geometric error. Degrees of freedom were 5.

No. steps velocity for acceleration estimated over	Time for velocity calculations used for acceleration (s)	R^2	F	Significance (p)
5	0.090	0.000	0.22	0.659
13	0.234	0.000	0.16	0.706
27	0.486	0.000	0.23	0.652
55	0.990	0.002	0.50	0.511
111	1.998	0.013	1.08	0.346

The x, y, z coordinates of the attitude directional vector with length 1 were calculated from roll (ϕ), pitch (θ) and heading (ψ):

$$z = \cos\theta\cos\phi \qquad 3.4$$

$$x = \sqrt{(1-z^2)}\cos(\operatorname{atan}(\frac{\sin\phi}{\sqrt{1-(\cos\theta\cos\phi)^2-(\sin\phi)^2}}+\psi) \qquad 3.5$$

$$y = \sqrt{(1-z^2)}\sin(\text{atan}(\frac{\sin\phi}{\sqrt{1-(\cos\theta\cos\phi)^2-(\sin\phi)^2}}+\psi) \quad 3.6$$

The distance r between the vectors at t_1 and t_2 is:

$$r = \sqrt{(x_{t1}-x_{t2})^2+(y_{t1}-y_{t2})^2+(z_{t1}-z_{t2})^2} \quad 3.7$$

And the angular velocity (v_{ang}) is:

$$v_{ang} = \frac{2\text{asin}(r/2)}{(t_2-t_1)} \quad 3.8$$

It can be assumed that angular acceleration is the difference between vectors describing the attitude between t_1 and t_2 and between t_2 and t_3 where t_1 - t_2 and t_2 - t_3 are equal.

The distance q between the ends of the vectors is:

$$q = \sqrt{(x_{t1}-2x_{t2}+x_{t3})^2+(y_{t1}-2y_{t2}+y_{t3})^2+(z_{t1}-2z_{t2}+z_{t3})^2} \quad 3.9$$

And the angular acceleration (a_{ang}) is:

$$a_{ang} = \frac{2\text{asin}(q/2)}{(t_2-t_1)^2} \quad 3.10$$

This calculation assumes that t_1 - t_2 and t_2 - t_3 are equal.

When angular velocity was calculated using consecutive values (18 ms) the results were noisy (Figure 3.15). When calculating the second derivative, angular acceleration, this noise was increased, resulting in the acceleration value calculated being mostly noise (Figure 3.16). To minimise this effect, velocity values were averaged over 1, 3, 5, 7 and 9 steps. This reduced noise, particularly in the acceleration values (Figure 3.17).

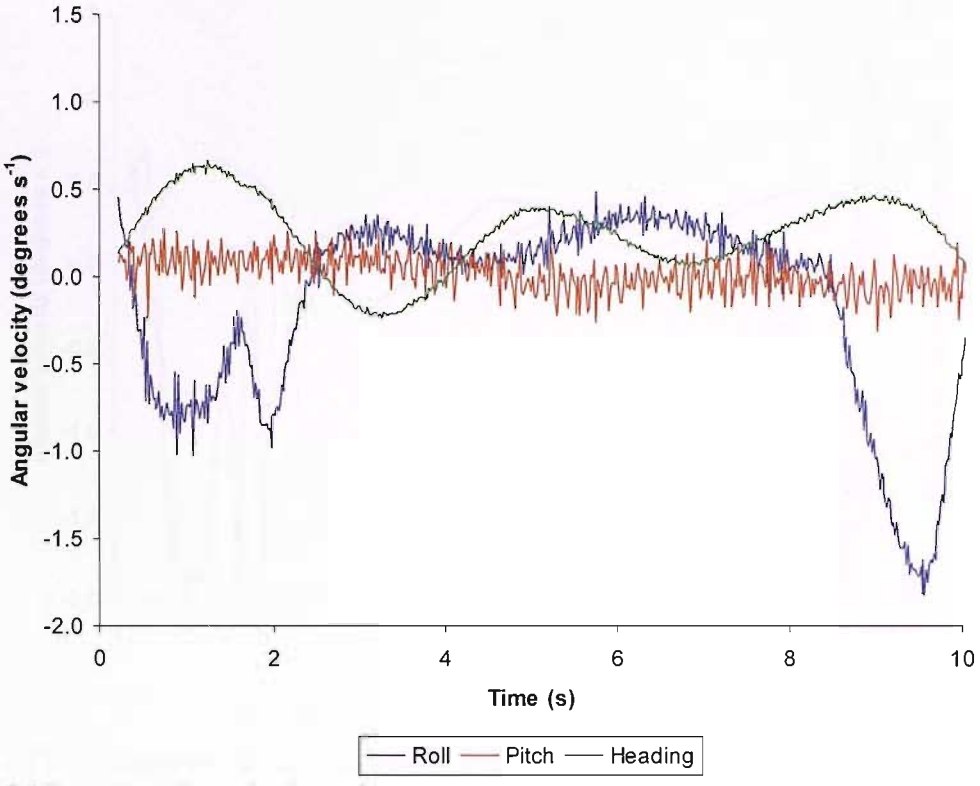


Figure 3.15 Angular velocity values calculated over 1 step or 18 ms.

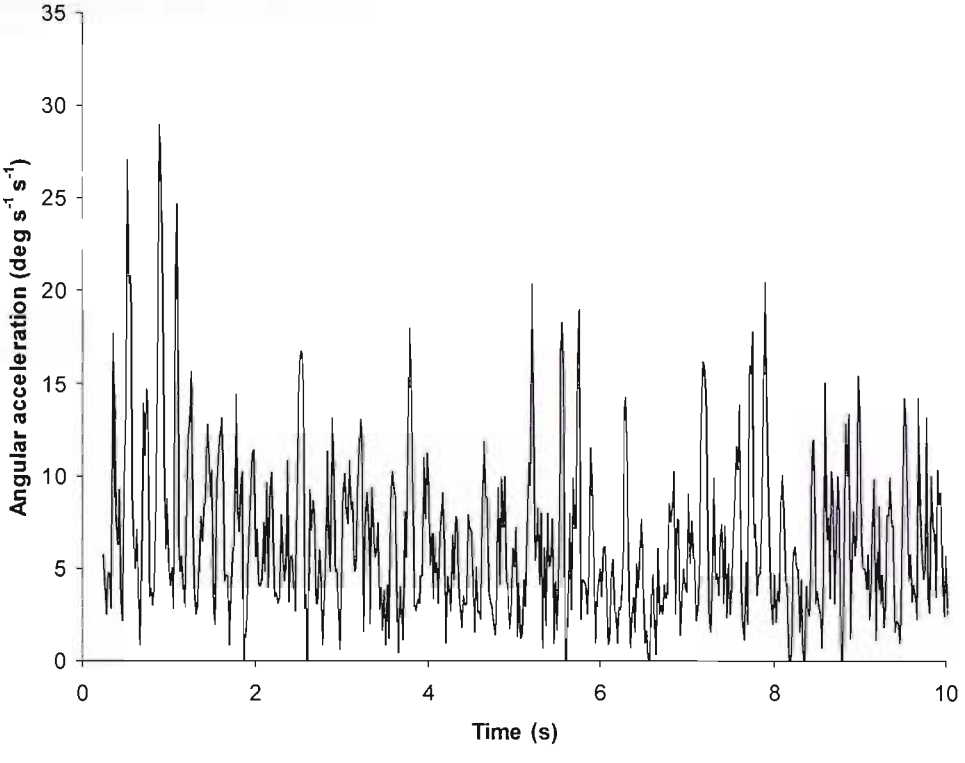


Figure 3.16 Angular acceleration values calculated over 1 step or 18 ms.

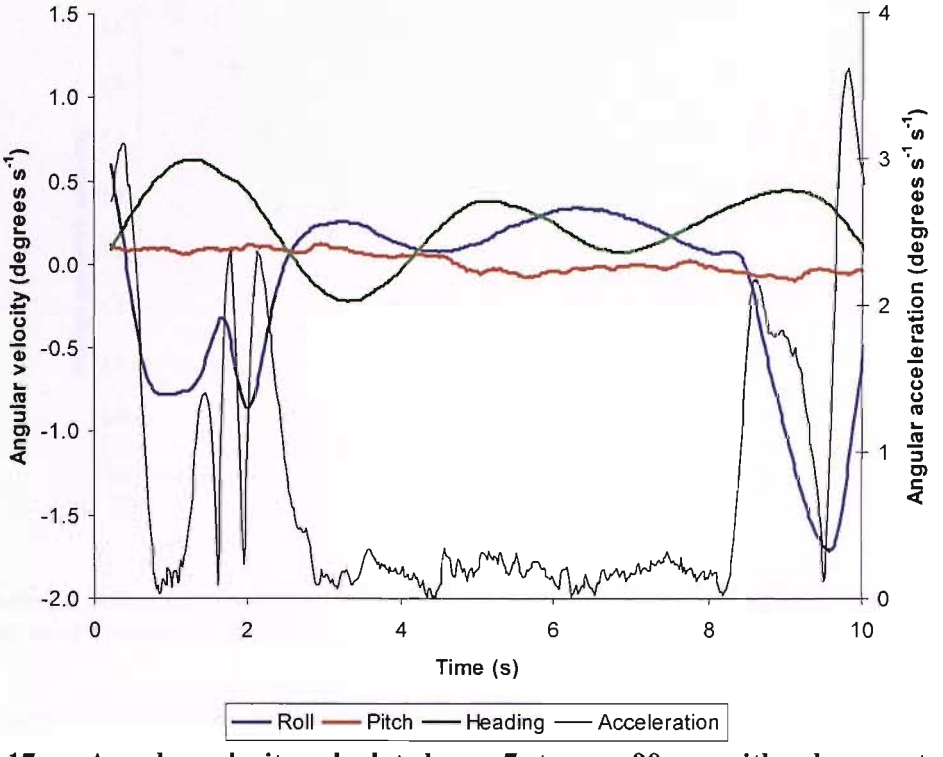


Figure 3.17 Angular velocity calculated over 7 steps or 90 ms, with subsequent acceleration values.

Angular velocity was not significantly correlated to geometric error (at 95% confidence) in all regressions carried out (Table 3.3; Figure 3.18).

Table 3.3 Linear regression of average angular velocity against average geometric error. Degrees of freedom were 5.

No steps angular velocity averaged over	Time velocity averaged over (ms)	R^2	F	Significance (p)
1	18	0.120	1.82	0.236
3	54	0.244	2.93	0.147
5	90	0.126	1.87	0.230
7	126	0.117	1.79	0.238
9	162	0.078	1.51	0.274

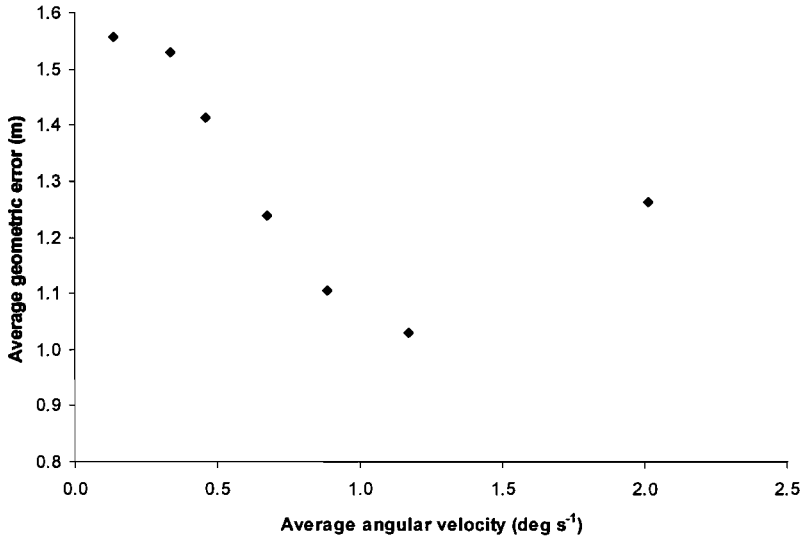


Figure 3.18 Geometric error as a function of angular velocity. Angular velocity was calculated over 7 steps or 90 ms.

Average geometric error was correlated with acceleration (at 95% confidence) when the velocity was averaged over 90 ms (Table 3.4; Figure 3.19).

The linear regression of geometric error as a function of angular acceleration resulted in the following equation:

$$E = 0.0742\ a_{ang} + 1.1661 \qquad 3.11$$

where E is the geometric error.

Table 3.4 Linear regression of average angular acceleration against average geometric error. Degrees of freedom were 5.

No. steps angular acceleration averaged over	Time that velocity used in acceleration calculation averaged over (ms)	R^2	F	Significance (p)
3	18	0.000	0.92	0.381
5	54	0.064	1.41	0.289
7	90	0.651	12.21	0.017
9	126	0.377	4.63	0.084

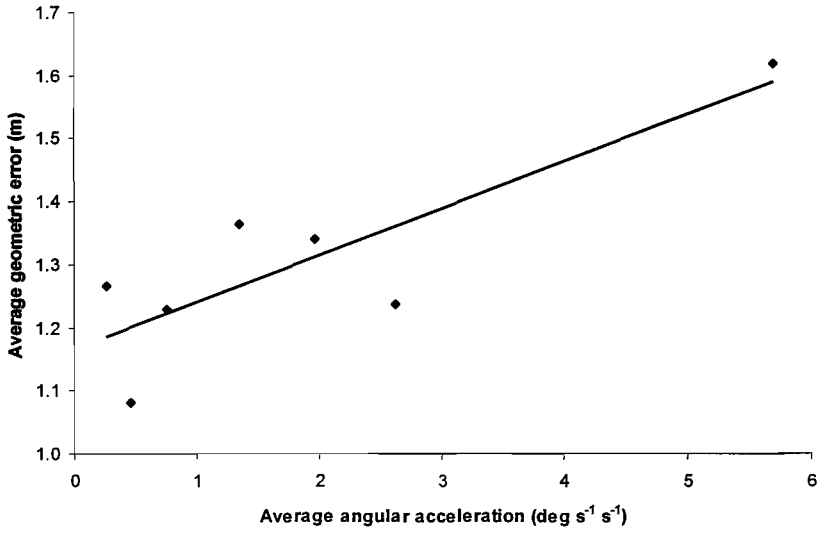


Figure 3.19 Geometric error as a function of angular acceleration. Acceleration averaged over 7 steps or 90 ms. Linear regression line. R^2=0.651; F=12.2; p = 0.017.

3.6.4 Local model of CASI instrument errors

The histogram in Figure 3.8 was estimated globally using all acceleration values. To account for the effect of angular acceleration, each error value obtained in the test phase was used to derive a geometric error distribution that was dependent on angular acceleration and flying height. To do this for a given pixel, the original geometric error ($E_{original}$) was modified using Equation 2.11, the angular acceleration ($a_{ang\,original}$). A function was applied that normalised the equation for the height above ground (z_{AGL}) as a proportion of the z_{AGL} assumed when Equation 2.11 was derived (900 m) to give an acceleration dependent geometric error ($E_{modified}$):

$$E_{modified} = \frac{E_{original}(0.0742\ a_{ang\,modified} + 1.1661)}{(0.0742\ a_{ang\,original} + 1.1661)} \cdot \frac{z_{AGL}}{900} \qquad 3.12$$

where $a_{ang\,modified}$ is the angular acceleration the distribution is being calculated for.

The modified geometric errors were combined to give an error distribution for angular acceleration $a_{ang\,modified}$ as in Figure 3.20 where error distribution has been calculated for two acceleration values and z_{AGL}= 900 m. Increasing the acceleration increases the maximum error values and mode and flattens the distribution curve.

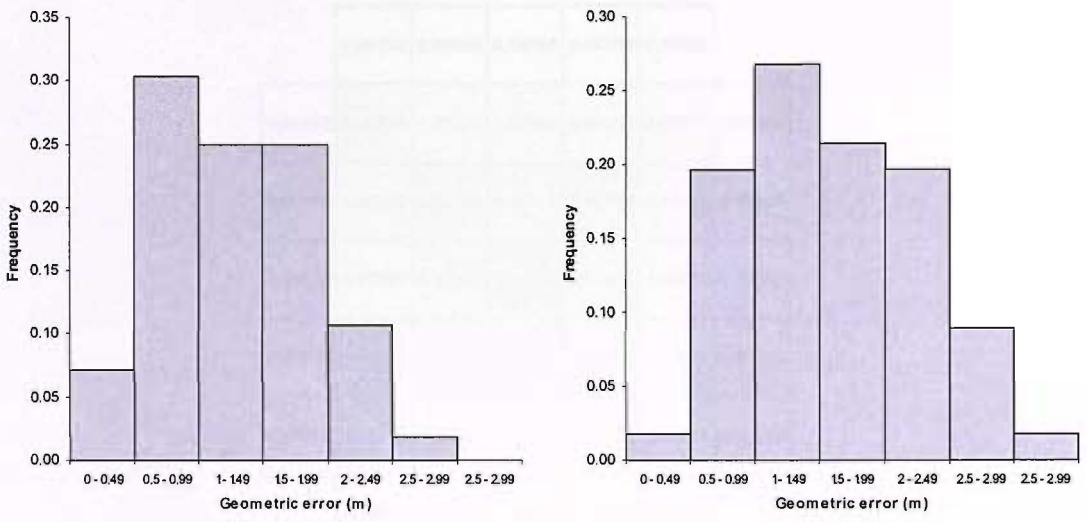


Figure 3.20 Effect of acceleration on predicted frequency of CASI instrument geometric errors.
a) Acceleration 1° s^{-2}
b) Acceleration 5° s^{-2}

A geometric error matrix with the same spatial resolution as the CASI imagery (1 m) was then derived (Figure 3.21). The matrix described CASI instrument error in terms of the probability that the pixel would occupy a given position within the matrix. The size of the matrix was dependent on the geometric error histogram. For each pixel in the matrix the probability of the geometric error being within that pixel (p_{Ej}) was calculated using:

$$p_{Ej} = \sum_{b=d_{\min}}^{d_{\max}} f_{b/b+0.5}(A_i/(\pi(b+0.5)^2 - \pi\, b^2)) \qquad 3.13$$

Where $f_{b\,b+0.5}$ is the frequency geometric errors with a magnitude between b and b+0.5. This was the derived angular acceleration dependent distribution (Figure 3.20).

d_{min} is the minimum distance of the pixel from the centre of the matrix

d_{max} is the maximum distance of the pixel from the centre of the matrix

A_j is area of the pixel between b and b+0.5 from the centre of the matrix

This was used to generate a matrix in which the probability of geometric error associated with every pixel in the filter was estimated (Figure 3.21). The matrix was a local measure of CASI instrument geometric uncertainty, as it was recalculated for every pixel and was dependent on angular acceleration and height above ground.

	0.00002	0.00084	0.00065	0.00084	0.00002	
0.00002	0.00200	0.02535	0.02095	0.02535	0.00200	0.00002
0.00084	0.02535	0.05500	0.09522	0.05500	0.02535	0.00084
0.00065	0.02095	0.09522	0.09501	0.09522	0.02095	0.00065
0.00084	0.02535	0.05500	0.09522	0.05500	0.02535	0.00084
0.00002	0.00200	0.02535	0.02095	0.02535	0.00200	0.00002
	0.00002	0.00084	0.00065	0.00084	0.00002	

Figure 3.21 Probability of CASI instrument geometric error filter generated for an angular acceleration of 1 deg s^{-1} s^{-1}. The number within each block of the matrix represents the probability that the pixel that should have occupied the centre of the matrix actually occupied that position due to instrument geometric error. The size of the matrix is dependent on the predicted geometric error distribution as in Figure 3.20.

The final CASI instrument geometric uncertainty model consisted of a geometric probability filter that adjusted for changes in acceleration and height of the sensor by expanding and contracting the distribution of geometric errors.

3.6.5 Orthometric error model

The CASI instrument error model was combined with the orthometric model, to provide an overall geometric error model. Orthometric errors are taken into consideration using a simple model based on the viewing geometry of the CASI instrument combined with the DEM used in geocorrection. Two viewing geometry angles were derived from the airborne post-processed GPS:

1. Off nadir angle. The angle between nadir (i.e. vertically below the sensor) and the look direction of a pixel. Look direction being the direction of a pixel from the sensor.
2. Azimuth angle. The horizontal angle between true north and the look direction of a pixel.

For every pixel in the geometric error filter (Figure 3.21), the difference between the elevation at that pixel derived from the DEM (z_{DEM}) and the elevation used in the CASI automated

geocorrection process (z_{ortho}) provided the potential elevation error. The off nadir angle (γ) and the elevation error of the pixel were used to derive the magnitude of the orthometric error (o) towards the CASI instrument using:

$$o = \tan(\gamma)(z_{ortho} - z_{DEM}) \quad 3.14$$

The azimuth angle (α) and o were then used to estimate direction of orthometric error in x and y:

$$x_{ortho} = \sin(\alpha) * o \quad 3.15$$

$$y_{ortho} = \cos(\alpha) * o \quad 3.16$$

The orthometric error in x and y were combined with the assumed error for each pixel in the matrix to provide an overall error vector. The predicted orthometric error was calculated using code written in ERDAS Macro Language (Appendix E). Each error vector had a probability associated with it derived from the matrix (Figure 3.21). The final local geometric uncertainty model was a series of vectors with probabilities derived from a combination of the instrument uncertainty and orthometric errors (Figure 3.22).

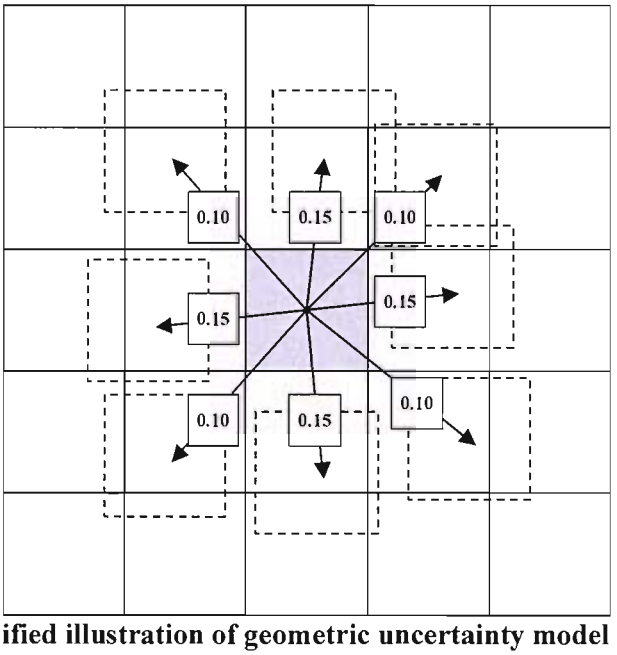


Figure 3.22 Simplified illustration of geometric uncertainty model for one pixel. Arrows represent geometric error vectors. The numbers represent the probability of the vector occurring. The squares with the dashed line represent a possible location for the pixel in the centre. The vectors are not whole pixels in x and y as each vector accounts for orthometric errors, which will not be exactly one pixel.

3.6.6 Misregistration model

To test the geometric error model, many easily recognisable points on a variety of slopes would have needed to be surveyed. This was not practical for this study as additional survey time was not available and so an alterative approach was taken. The geometric models for two CASI images were combined to generate a misregistration model. Points on both automatically geocorrected images could be identified and at these points the actual and model predicted misregistration could be compared without the need for ground surveying. The potential errors within each pixel in the matrix for 2001 were combined with the potential errors within each pixel in the matrix for 2002. The combined probabilities were used to predict a mean misregistration error.

$$x_{error} = (x_{ortho,n} - x_{ortho,m}) + (x_{pixel,n} - x_{pixel,m}) \quad 3.17$$

$$y_{error} = (y_{ortho,n} - y_{ortho,m}) + (y_{pixel,n} - y_{pixel,m}) \quad 3.18$$

Where x_{ortho} is the error in x due to orthometric effects

x_{pixel} is the x offset from the central pixel of the matrix (Figure 3.21)

n is a pixel in matrix t_1

m is a pixel in matrix t_2

The misregistration error between two pixels in matrix t_1 and matrix t_2 ($\mu_{n,m}$) is given by:

$$\mu_{n,m} = (x_{error}{}^2 + y_{errror}{}^2)^{0.5} \quad 3.19$$

and overall misregistration μ is given by:

$$\mu = \frac{\sum_{n=1}^{\varepsilon_{t1}} \sum_{m=1}^{\varepsilon_{t2}} \mu_{n,m} \rho_n \rho_m}{\varepsilon_{t1} \varepsilon_{t2}} \quad 3.20$$

where ε_{t1} is the number of pixels in the t_1 CASI instrument geometric uncertainty matrix

ρ_n and ρ_m are the probability of geometric error for matrix t_1 and matrix t_2 respectively. They were derived from Figure 3.21.

3.6.7 Testing misregistration models

CASI and LiDAR data acquired on 28th August 2001 and 11th September 2002 (Section 2.6.1) over the urban area to the north-east of the Ainsdale test site were used to test the misregistration models (Figure 2.2; Figure 3.23). The positions of 140 easily identifiable points on the 2001 and 2002 datasets were measured (Figure 3.23) and the actual misregistration measured between the images was compared to that estimated by the model.

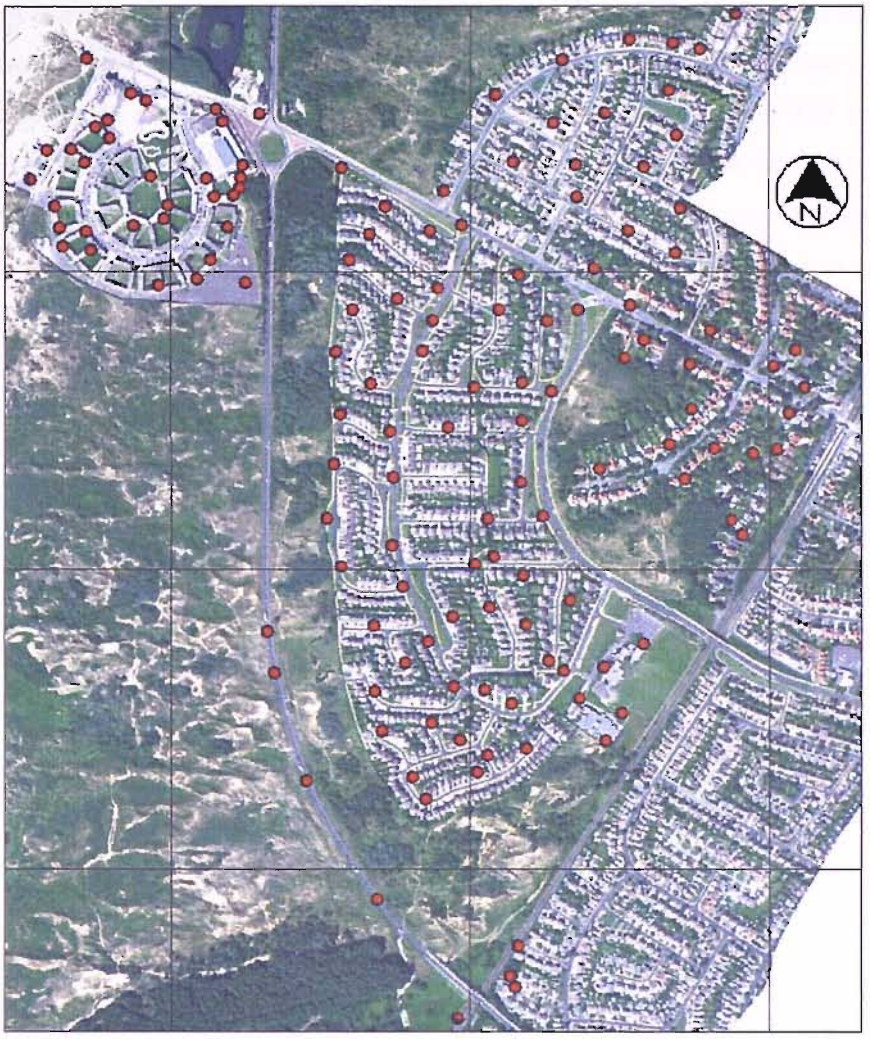


Figure 3.23 CASI data of Ainsdale urban misregistration test site with test points overlaid as red points. Grid is 500m.

To account for some of the errors within the LiDAR DSMs the geometric models were tested using the following DSM models:

1. Nearest neighbour DSM: the original 2 m LiDAR DSM was resampled to the same 1 m grid as the CASI data using nearest neighbour resampling. This model did not account for the uncertainty associated with the LiDAR data.

2. Cubic convolution DSM: the original 2 m LiDAR DSM was resampled to a 1 m grid using cubic convolution resampling. This model provided a potentially more realistic model of the elevation surface, as values vary between adjacent pixels, something that did not happen using nearest neighbour resampling.
3. LiDAR uncertainty: the model was run using the elevation values from the DSM in 1 above. The quoted errors in the LiDAR system (Ashkenazi, 1999) were used to account for uncertainty in the LiDAR DSM. The predicted RMSE errors of the LiDAR are 0.8 m in the horizontal and 0.25 m in the vertical. Uncertainty in the vertical was accounted for by estimating misregistration error using three DSM surfaces, the original surface (DSM A), one with 0.25 m subtracted (DSM B) and one with 0.25 m added (DSM C). To maintain a simple method of estimating misregistration, horizontal error was accounted for using a 1 m error rather than 0.8 m. To generate a simple model of LiDAR uncertainty the model was run with each of the three DSMs to account for vertical errors. To account for horizontal errors the nine DSM values within a three by three filter were used (Figure 3.24). This resulted in the geometric error being estimated using 27 z values. Misregistration was estimated using:

$$\mu = \frac{\sum_{DSM=1}^{9C} \sum_{n=1}^{\varepsilon_{t1}} \sum_{m=1}^{\varepsilon_{t2}} \mu_{n,m} \rho_n \rho_m}{\varepsilon_{t1} \varepsilon_{t2}} \qquad 3.21$$

1	2	3
4	5	6
7	8	9

Figure 3.24 Matrix used in LiDAR uncertainty model.

Each of the above DSM models was run with and without angular acceleration, resulting in six different models. Without angular acceleration the geometric uncertainty model was a global CASI instrument error model combined with a local orthometric model. When the angular acceleration was used, the geometric uncertainty model was a local CASI instrument model combined with a local orthometric model.

To compare the actual and predicted misregistration, predicted misregistration was sorted by magnitude of error for every point and binned in groups of ten to generate averaged predicted and averaged actual misregistration errors. Linear regressions were carried out on the averaged predicted and averaged actual misregistration errors to determine whether there was a significant correlation (Table 3.5). An *F*-test was carried out assuming that the predicted misregistration error should equal the actual misregistration error (Table 3.6).

Table 3.5 Regression of predicted versus actual misregistration. Degrees of freedom were 12.

Acceleration used	Model used	R^2	Constant	Slope	*F*-value	Significance (*p*)
Yes	Nearest neighbour DSM	0.389	-0.20	1.09	9.27	0.010
No	Nearest neighbour DSM	0.560	-2.06	2.10	17.52	0.001
Yes	Cubic convolution DSM	0.275	0.22	0.85	5.93	0.031
No	Cubic convolution DSM	0.315	-1.80	1.92	6.98	0.021
Yes	LiDAR uncertainty	0.208	-0.34	1.15	4.41	0.057
No	LiDAR uncertainty	0.060	-0.84	1.43	1.84	0.200

Table 3.6 *F*-test assuming predicted misregistration equals actual misregistration. Degrees of freedom were 12.

Acceleration used	Model used	*F*-value	Significance (*p*)
Yes	Nearest neighbour DSM	7.59	0.017
No	Nearest neighbour DSM	4.09	0.066
Yes	Cubic convolution DSM	7.06	0.021
No	Cubic convolution DSM	2.97	0.110
Yes	LiDAR uncertainty	3.23	0.098
No	LiDAR uncertainty	0.89	0.365

3.6.7.1 Nearest neighbour LiDAR DSM

Using the nearest neighbour LiDAR DSM in the geometric error model, a statistically significant linear correlation (at 95% confidence) was found between predicted misregistration and actual misregistration for both the model with angular acceleration and the model without (Table 3.5; Figure 3.25; Figure 3.26). An *F*-test carried out assuming predicted misregistration should equal actual misregistration was not significant (at 95% confidence) when angular acceleration was not used in the geometric model (Table 3.6). When angular acceleration was used the *F*-test was significant (at 95% confidence) (Table 3.6).

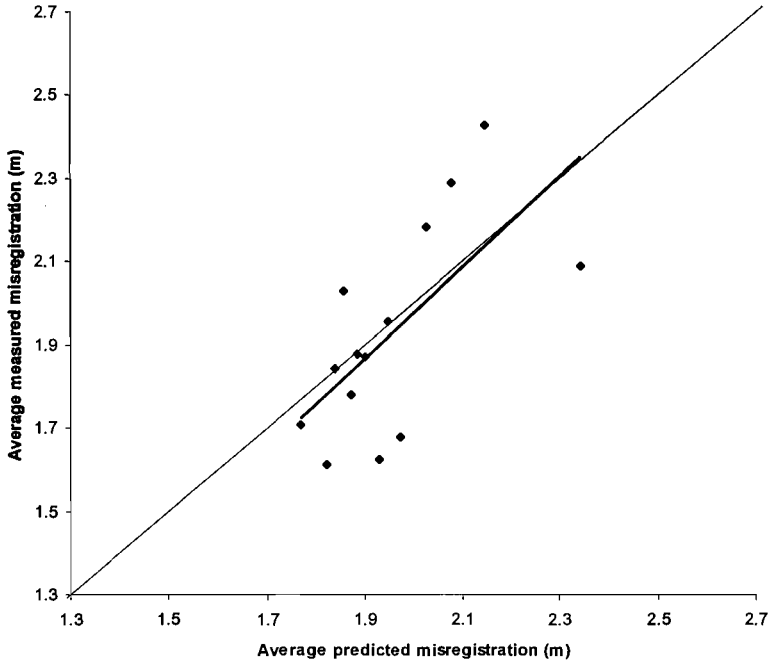


Figure 3.25 Predicted versus actual misregistration not using acceleration in misregistration model and the 2m LiDAR DSM. Dark line is regression line. Thin line represents actual misregistration equals predicted misregistration.

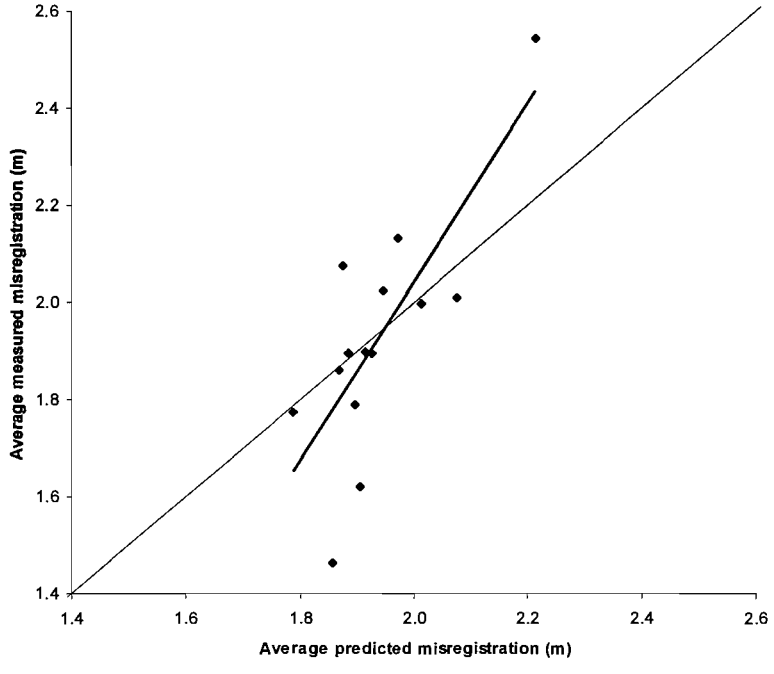


Figure 3.26 Predicted versus actual misregistration using acceleration in misregistration model and the 2 m LiDAR DSM. Dark line is regression line. Thin line represents actual misregistration equals predicted misregistration.

This shows that though there was a correlation between predicted and actual misregistration for both models using the 2 m LiDAR DSM, when angular acceleration was not used, the model

did not predict misregistration accurately. The optimum relationship between predicted and actual misregistration would be one in which the slope was close to 1 and the offset was close to 0. The regression when angular acceleration was used resulted in an offset of 0.20 and a slope of 1.09 (Figure 3.26), compared to -2.06 and 2.10 respectively for the regression when angular acceleration was not used (Figure 3.25).

3.6.7.2 Cubic convolution LiDAR DSM

Using the cubic convolution LiDAR DSM in the geometric error model, a statistically significant correlation was found between predicted misregistration and actual misregistration for the model using angular acceleration and the model that did not use angular acceleration (Table 3.5; Figure 3.27; Figure 3.28). Again, an *F*-test carried out assuming that the predicted misregistration should equal actual misregistration was not significant (at 95% confidence) when angular acceleration was not used in the geometric model (Table 3.6).

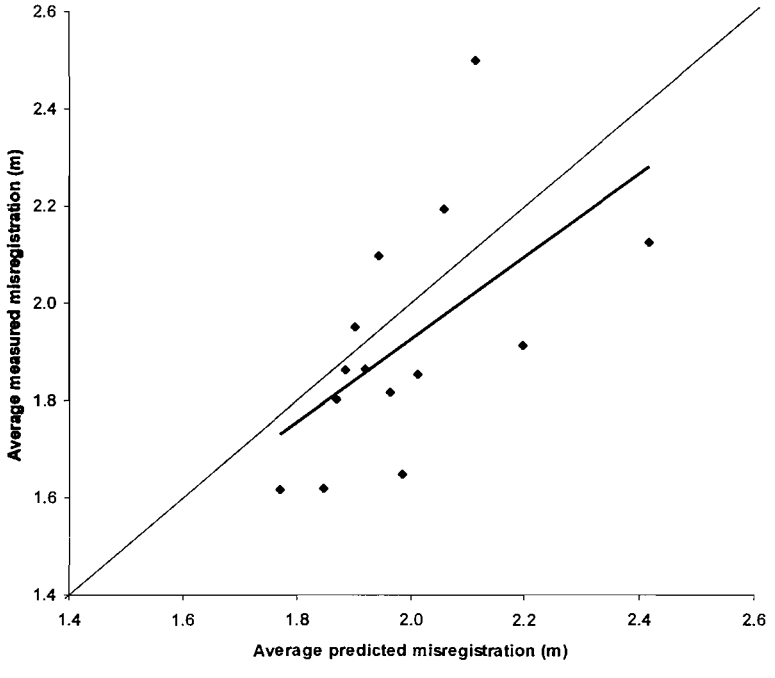


Figure 3.27 Predicted versus actual misregistration using the cubic convolution LiDAR DSM in misregistration model but not using angular acceleration. Dark line is regression line. Thin line represents actual misregistration equals predicted misregistration.

When angular acceleration was used the *F*-test was significant (Table 3.6). There was a significant correlation between predicted and actual misregistration for both models using the cubic convolution LiDAR DSM. However, when angular acceleration was not used, the model did not predict misregistration accurately, as the regression offset was -1.80 and the slope 1.92

(Table 3.5; Figure 3.27). The regression when angular acceleration was used resulted in an offset of 0.22 and a slope of 0.85 (Figure 3.28), indicating that this model was not as accurate as the model using the nearest neighbour LiDAR DSM.

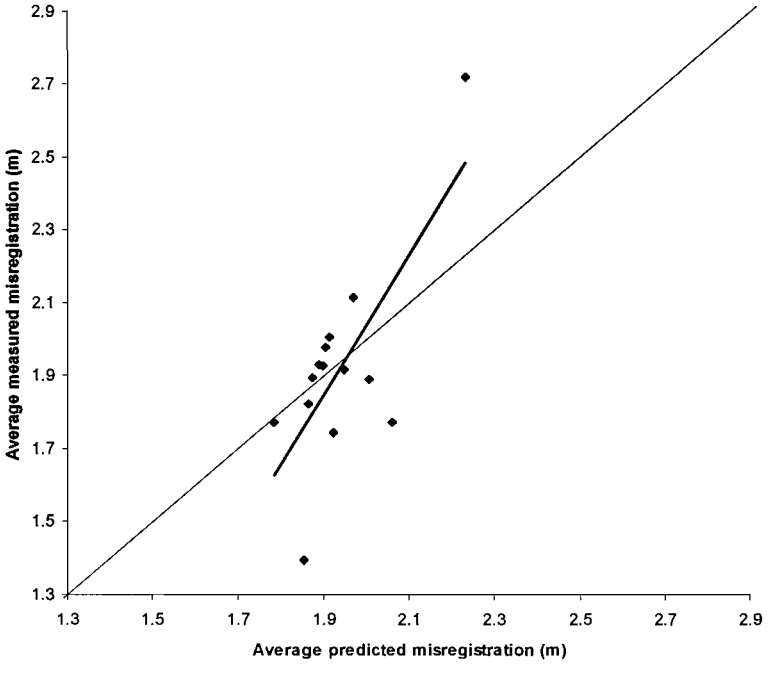


Figure 3.28 Predicted versus actual misregistration using acceleration in misregistration model and the cubic convolution LiDAR DSM. Dark line is regression line. Thin line represents actual misregistration equals predicted misregistration.

3.6.7.3 LiDAR uncertainty model

Using the LiDAR uncertainty misregistration model a statistically significant correlation was not found between predicted and actual misregistration for the models that did and did not use angular acceleration (at 95% confidence) (Table 3.5; Figure 3.29; Figure 3.30). *F*-tests carried out assuming that the predicted misregistration should equal actual misregistration were not significant for both models (at 95% confidence) (Table 3.6). These results indicated that these models were the least accurate of those tested.

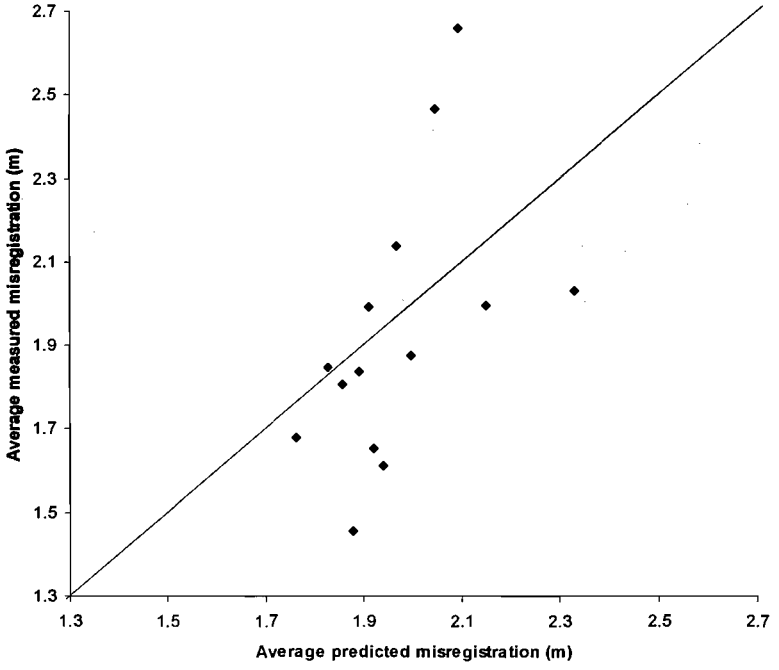


Figure 3.29 Predicted versus actual misregistration using the LiDAR uncertainty misregistration model but not using angular acceleration. Thin line represents actual misregistration equals predicted misregistration.

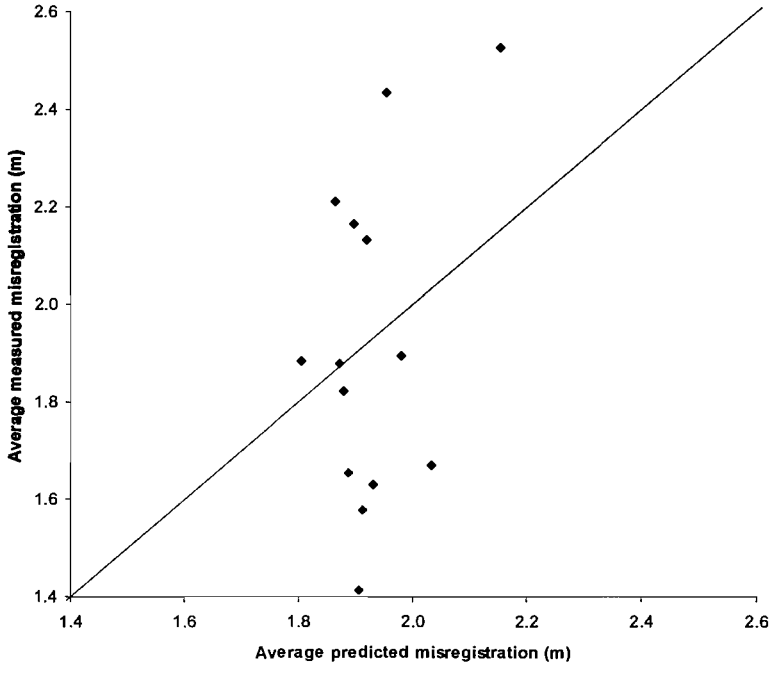


Figure 3.30 Predicted versus actual misregistration using acceleration in the LiDAR uncertainty misregistration model. Thin line represents actual misregistration equals predicted misregistration.

3.7 Discussion and conclusions

3.7.1 CASI instrument error

This study examined the magnitude of CASI instrument geometric errors in order to derive a probabilistic model of those errors. The correlation between error and a variety of variables was tested, but a significant correlation was found only with angular acceleration (R^2=0.651; F=12.2; p=0.017) (Table 3.1; Table 3.2; Table 3.3; Table 3.4). The 63 points used for testing geometric uncertainty were averaged in seven bins. Though there could have been correlations between the magnitude of geometric error and other variables, the sample size in this study may have been too small to find significance for some of the variables. If this were the case then more precise models of geometric error could be derived for the CASI system using additional variables if a larger sample size was used. Another limitation is that the variables were likely to have a small effect on geometric error, but the precision of the measurements was one pixel or 1 m.

A matrix based model of CASI instrument error was generated using a global measure of uncertainty derived from the histogram of geometric errors and the empirically derived regression model of the geometric error as a function of angular acceleration (Figure 3.21). The probability of error associated with each pixel in the matrix was derived for every pixel in an image. Every pixel within a CASI image had a matrix associated with it that provided the local geometric error distribution due to CASI instrument errors.

3.7.2 Overall CASI geometric error

The matrix based CASI instrument uncertainty model was combined with an orthometric error model based on a LiDAR DSM data to provide a model of CASI geometric uncertainty. The model combined instrument errors and terrain effects in a probabilistic model of geometric error. This modelled the spatial variation of geometric errors in CASI imagery. Though it was not possible within this study to test the CASI geocorrection model directly, as ground data were not available, it was possible to use the model to generate a misregistration model. By combining the geometric uncertainty models for two sets of CASI imagery, it was possible to combine the probabilistic models of geometric uncertainty into a single averaged prediction of misregistration that could be tested against measured values.

Six misregistration models were tested that incorporated three different LiDAR DSM models. Each of the LiDAR DSM models was tested with and without angular acceleration. Of the six

models tested there was a significant relationship (at 95% confidence) between predicted and actual misregistration for two.

The use of angular acceleration in the geometric uncertainty models resulted in a smaller correlation between predicted and actual misregistration for the Nearest Neighbour and Cubic Convolution models (Table 3.5). However, for the model to be accurate, predicted misregistration should equal actual misregistration, and *F*-tests showed that this equal relationship was only significant when angular acceleration was used (Table 3.6). This indicated that use of angular acceleration produced more accurate, but less precise models of geometric uncertainty compared to not using angular acceleration. The reduction in precision when angular acceleration was used may indicate that the relationship between geometric error and angular acceleration was not linear, the relationship was noisy or that other variables influenced geometric error.

3.7.3 Effect of DSM on geometric error prediction

When angular acceleration was used in the misregistration model, the nearest neighbour DSM method had the largest correlation between predicted and actual misregistration ($R^2 = 0.389$; p=0.010) of the three DSM inputs (Table 3.5). The use of the cubic convolution DSM resulted in much smaller but still statistically significant correlation ($R^2 = 0.275$; p=0.031). The LiDAR uncertainty model had the smallest correlation and this was not significant ($R^2 = 0.208$; p=0.057).

F-tests assuming a one-to-one relationship between predicted and actual misregistration resulted in the largest *F*-test value for the model using the nearest neighbour DSM (Table 3.6). Of all the models tested, the one that used angular acceleration and the nearest neighbour DSM was the most accurate, with the least difference from the optimum where the offset equalled 0 and the slope equalled 1 from the regression of predicted against actual misregistration (Table 3.5). This means that the model using the nearest neighbour DSM resulted in the most accurate, statistically significant, one-to-one representation of misregistration.

The model used in this study to incorporate LiDAR uncertainty in a prediction of misregistration was the least accurate method of predicting misregistration and the one-to-one relationship between predicted and actual misregistration was not statistically significant (at 95% confidence).

The model used was simplistic and a more accurate and possibly more complex model of LiDAR uncertainty is required. To include a more accurate model of LiDAR uncertainty, a fuller understanding of the errors and uncertainties within the data is needed. A three-dimensional error probability distribution may be required, similar to the two-dimensional CASI geometric error distribution derived in this study. However, determining LiDAR errors in three dimensions simultaneously may be difficult. Errors in the vertical may be estimated by comparing height from ground survey with height from LiDAR on a flat surface. Horizontal positional errors can be determined where there is a vertical structure, such as the side of a building. Determining vertical and horizontal errors simultaneously may be more difficult. On flat surfaces, it is not possible to determine horizontal errors from elevation. On a vertical structure, it is not possible to determine vertical error. It may be possible to use the LiDAR intensity image, an additional product that provides an estimate of backscatter to determine a three-dimensional LiDAR error model. If a survey were carried out over a flat surface that had markings that would show up in the wavelength of the LiDAR used, it would be possible to simultaneously measure horizontal and vertical errors using the intensity image. Ground survey elevation could then be compared to LiDAR elevation and the intensity image could be used to determine horizontal errors. The correlation between these errors and variables such as measures of GPS quality and variables that effect IMU accuracy could be determined.

3.7.4 Modelling misregistration in change detection

The overall misregistration measure, μ, used in this chapter was an average value and could be used to model misregistration using an approach such as eroding the boundaries between classes (Veregin, 1996; Serra *et al.*, 2003). However, though this approach can increase accuracy of change detection (Serra *et al.*, 2003), there are limitations. Using μ did not account for the range of possible misregistration errors, as the range of possible misregistration errors were combined to provide one value.

As μ was a single value per-pixel, there were no probabilities associated with the misregistration and no directional component to the misregistration, making μ unsuitable as a local measure of misregistration uncertainty and therefore it was not used in the final change detection model.

The geometric error vector and associated probability were potentially more suitable for use in a change detection uncertainty model than μ. The probability values associated with each geometric uncertainty vector may be combined with probabilistic measures of thematic

uncertainty to provide an overall probabilistic measure of the uncertainty associated with change detection.

3.7.5 Increasing model accuracy

Though studies have attempted to quantify errors for a variety of IMU models (Grejner-Brzezinska *et al*, 1998; Hong *et al.*, 2002; Mostafa *et al.*, 2001), these have not discussed the variation in errors or tried to find the relationships with variables such as angular acceleration. Future work should be carried out that provides an increased understanding of these errors so that geometric uncertainty can be understood and predicted with greater accuracy.

The work carried out in this study used raw IMU and post-processed GPS navigational data. Errors in navigational data can be reduced by merging the attitude and positional data to derive a post-processed navigational solution (Mostafa and Hutton, 2001; Toth, 2002). This approach integrates GPS and IMU data to provide a navigational solution that has reduced errors. However, the effect of acceleration on the accuracy of the post-processed navigational solution is complex, as accelerations can reduce errors (Mostafa and Hutton, 2001). This means that the input variables used in this study may not be suitable for input when using post-processed, integrated GPS and IMU data.

3.7.6 Applying misregistration model to other platforms

The principles of geometric error model generation for the CASI instrument could be used for other airborne systems. Airborne sensors such as HyMap (Kruse *et al.*, 2000), AISA+ (Specim, 2004), AVIRIS (Boardman, 1999) can use post-processed GPS and IMU based systems to provide navigational data for automatic geocorrection. The principles of generating a geometric error model used in this study would be directly applicable to other instruments.

Though satellite systems are not exposed to the sudden alterations in motion that occur on an airborne platform, some of the methods used in the CASI model would be applicable, particularly the use of topographic data. It may be possible to derive a model of instrument error for satellite based systems, though the variables may be different from airborne systems. Even if it was not possible to derive a local model of instrument uncertainty, a global model could be derived from the distribution of geometric errors. This could be combined with a local orthometric error model derived from a DEM to provide a local model of geometric errors. Though this model would not be as precise as one derived from a local model of instrument

geometric errors, it would have a local function and would compensate for the increase in geometric errors from satellite-based imagery with an increase in topographic variation (Stow, 1999). The Shuttle Radar Topography Mission (SRTM) in 2000 acquired data that would be suitable for this purpose. SRTM acquired elevation data for all land areas between 60° North and 56° South latitude, approximately 80% of the Earth's land mass (Rabus *et al.*, 2003). The data have a horizontal spatial resolution of approximately 30 m, with a vertical accuracy of 16 m at 90% confidence (Rabus *et al.*, 2003). This horizontal resolution is of a similar order to medium resolution satellite based data such as Systeme Pour l'Observation de la Terre (SPOT) High Resolution Visible (HRV) multispectral (20 m) (SPOT, 2005) and Landsat Thematic Mapper (TM) (30 m) (NASA, 2002) making it potentially suitable for these and coarser resolution satellite data such as MODIS (NASA, 2004).

3.7.7 Summary

In this chapter, probabilistic geometric error models were generated that modelled the spatial function of geometric error in CASI imagery, using angular acceleration of the sensor, a global model of the probability of CASI geometric errors and a simple orthometric error model. The geometric uncertainty model consisted of a series of geometric error vectors, each with a probability associated (Figure 3.22).

It was not possible to test the geometric uncertainty model and so a misregistration model that provided an average misregistration error was derived from the geometric error. This misregistration model accurately modelled the spatial function of misregistration error between two CASI mosaics. Though the misregistration model was not suitable for probabilistic change detection, the geometric uncertainty model it was derived from was suitable, as it provided error vectors with associated probabilities on a per-pixel basis.

4 Thematic uncertainty

4.1 Introduction

The purpose of the research in this chapter was to examine methods of deriving estimates of thematic error for use in operational change detection (Figure 4.1). Estimates were made of thematic error resulting from classifications of airborne sensor data and took the form of thematic uncertainty measures. The relationships between the variables used in the setup of classifiers and the accuracy of thematic uncertainty prediction were examined.

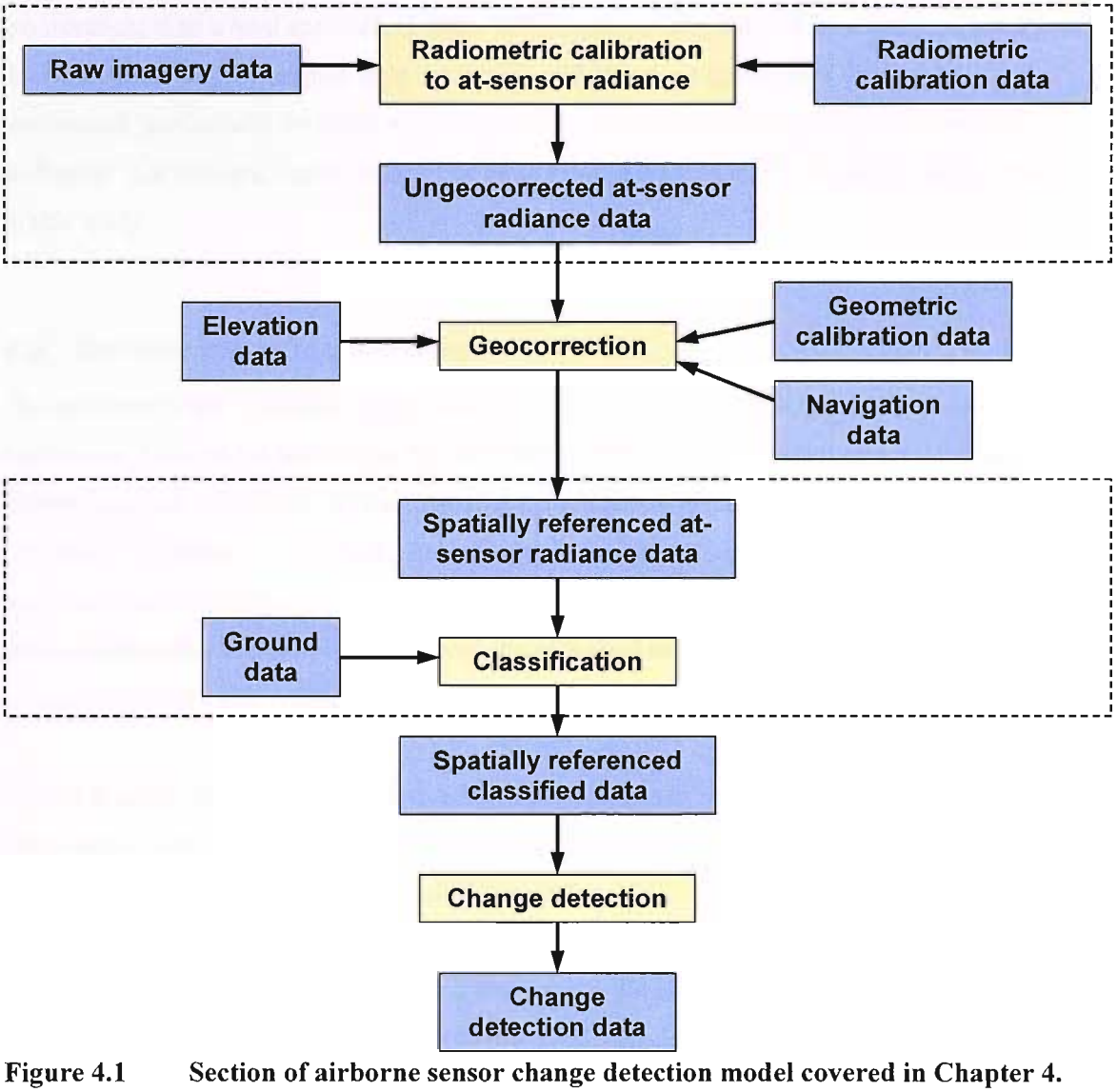


Figure 4.1 **Section of airborne sensor change detection model covered in Chapter 4.**

4.2 Hard and soft classifications

The purpose of this study was to examine methods of carrying out change detection using remote sensing for input into management plans and legislative reporting for the EA and EN. Potentially the change detection outputs will be examined by people without experience or understanding of geographical information science and remote sensing. In this context, relatively simple outputs that could be understood easily by both staff and stakeholders such as the public, industry and politicians were required. Two approaches to pixel based change detection could be taken, either hard or fuzzy. As multiple classes may exist within the area defined by a pixel, a fuzzy approach to classification may more accurately model the environment than a hard approach (Fisher, 1997; Arnot *et al.*, 2004). However, compared to a hard classification, the outputs from a fuzzy classification are much more difficult to understand, particularly for those with little or no geographical training, and are more difficult to display (Lucieer and Kraak, 2004). For these reasons a hard classification approach was taken in this study.

4.3 Deriving thematic uncertainty

Thematic errors are caused by factors such as errors in the image data due to inaccuracy in the calibration, noise in the sensor system, atmospheric effects, bi-directional effects such as shadowing, lack of spectral separability of classes, inaccurate training data or classifier errors (Figure 4.1) (Song and Woodcock, 2003; Carmel and Dean, 2004). Though these errors result in incorrect class allocation during the classification process, they are complex and outside the scope of this study. To simplify this study, the effects of sensor error, atmosphere and errors in ground data on thematic errors were assumed to be negligible and were not considered.

Thematic error measures may be derived on a per-class basis using the confusion matrix to estimate the number of pixels of a given class that are incorrectly allocated (Appendix B). The per-class error estimated from the confusion matrix could be used to derive a thematic uncertainty measure such as the probability of error on a per-class basis. However, it is unlikely that every pixel allocated to a given class has equal probability of membership of that class (Bishop, 1995). If a given pixel is close to the class centroid it is more likely to be classified correctly than a pixel that is further away from the class centroid (Bishop, 1995). A pixel that is made up of equal amounts of two classes is less likely to be correctly classified than a pure pixel, as there is likely to be confusion between the two classes.

Spatially unreferenced global thematic measures will not account for the spatial variation in probability of correct allocation. For pixels with a large probability of membership to a class, the per-class measure will overestimate probability of error. As the error is overestimated, the precision of change detection will be reduced. For pixels with a small probability of being correctly allocated to a class, a per-class measure will underestimate error. This will result in probability of error in change detection being underestimated, resulting in a larger probability of change detection error.

To determine where thematic errors are likely to occur, more information on per-pixel uncertainty is required than is available from the spatially unreferenced per-class measures that may be derived from the confusion matrix (Foody, 2005). The estimation of the probability of error on a pixel basis provides the opportunity to estimate more accurately the probability of error in change detection. Several studies have derived per-pixel measures of thematic uncertainty from classification outputs (Gong *et al.*, 1996; Shi and Ehlers, 1996; Ediriwickrema, and Khorram, 1997; de Bruin and Gorte, 2000; McIver and Friedl, 2001; Liu *et al.*, 2004).

In a per-pixel context for a hard classification, there are two measures that may be used to predict thematic uncertainty:

1. A measure that indicates whether the correct class has been allocated to a pixel. This requires definition of a single variable that indicates the probability of correct allocation for the class allocated to that pixel.
2. A series of measures that indicate whether any class would be correct if allocated to that pixel. This requires definition of variables that indicate the probability of correct allocation for every class.

4.4 Classifiers for deriving thematic uncertainty measures

4.4.1 Maximum likelihood classifier

The most common of the traditional parametric approaches to classification is maximum likelihood (ML) classifier (Schowengerdt, 1997; Campbell, 2002). The ML classifier provides an approach for classifying remotely sensed data that is relatively easy to understand and to carry out.

In the ML approach, class mean vectors in feature space and variance-covariance matrices for each class are generated from training data. Probability density functions for all classes are then derived from these statistics. The posterior probabilities of class membership are estimated for each pixel and the pixel is then allocated membership to the class that it has the greatest probability of membership to.

If the data approximate a normal distribution, this approach can produce accurate classifications (Benediktsson *et al.*, 1990). The analyst does not have to set any parameters for the classification to run, just select suitable training data and classes. To an analyst who has a minimal amount of statistical knowledge, the ML classifier is relatively simple to understand and most commercially available image processing packages (for example ERDAS Imagine, ENVI, IDRISI) have an easy to use ML classifier.

The ML classifier assumes the data approximate a multivariate Gaussian distribution (Campbell, 1981; Benediktsson *et al.*, 1990). This is often not the case, especially when a class contains a great deal of variation. For example, a woodland class may contain a variety of species, with mixtures of coniferous and deciduous species, resulting in a multi-modal distribution of the spectra.

The ML classifier is also very sensitive to the form and quality of ground data. There should be enough sampling points to represent the full variety within each class. According to Swain (1978) the amount of training data required for the ML classifier is linked to the dimensionality of the data set being classified. The greater the number of dimensions, the larger the training points sample size needs to be. If the training set is too small, ML classification accuracy may be reduced as dimensionality is increased (Lee and Langrebe, 1993). According to the study by Swain (1978), the sample size for each class should be at least 30 times and preferably 100 times the number of dimensions. Carrying out an ML classification using the 14 band bandset collected over the study sites (Table 2.2) it may be impractical to acquire the required number of training data for each class.

If the training data do not incorporate the variation within the classes used, the accuracy of the classification may be reduced, even if the required numbers of pixels are used for each class (Campbell, 1981). In many cases, training data are sampled in blocks of contiguous pixels. As these pixels may exhibit autocorrelation, the class statistics (means, variances and covariances), may inadequately represent the classes, leading to reduced classification accuracy (Campbell, 1981).

A requirement for statistical classification approaches is that the training data should only contain one land cover class (Paola and Schowengerdt; 1995). However, in the case of natural and semi-natural habitats this may be very difficult to achieve, as certain classes generally occur as mixtures with other classes.

The use of data from multiple sources has the potential to reduce ML classification accuracy (Peddle *et al.*, 1994). When multisource data are used in ML classification there are scaling issues that have to be considered, as the different data may not be in common units (Benediktsson *et al.*, 1990). Data sources are not equally reliable or useful in discriminating between classes and the ML classifier does not have a mechanism for weighting data according to importance (Benediktsson *et al.*, 1990). Thematic data may have the potential to increase class discrimination, but as they are non-parametric, they should not be used in ML classification, but if used can result in reduced classification accuracy.

Though the ML classifier is one of the most commonly used in remote sensing (Richards, 2005), in this study there are factors that make its use inappropriate. These factors include the use of remotely sensed data with large numbers of wavebands and small training data sets. As the number of training pixels for some classes, including Reed and Buckthorn, was small it is likely that the ML would have classified these classes inaccurately. Within some of the classes used in this study there were a variety of different vegetation types that could result in multi-modal distributions of spectral data. Using the ML classifier when data have multi-modal distributions would be likely to result in a reduced classification accuracy compared with non-parametric methods (Kanellopoulos *et al.*, 1992). Studies have shown that under the conditions described above there are non-parametric classifiers such as neural networks that can classify more accurately than the ML classifier (Kanellopoulos *et al.*, 1992; Peddle *et al.*, 1994; Yool, 1998).

4.4.2 Neural networks

Neural networks are the computing equivalent of a very simple biological brain. They provide a possible solution to a variety of problems in remote sensing including classification and biophysical property extraction.

Neural network and statistical methods of classification are fundamentally different in that statistical approaches depend on an assumed model, while neural networks depend on data

(Benediktsson *et al.*, 1993; Atkinson and Tatnall, 1997; Zhou, 1999). This means that the underlying assumptions made for statistical classification, such as the data are normally distributed and data layers are not correlated, do not need to be met for neural networks.

As well as being distribution free, neural networks are importance free (Benediktsson *et al.*, 1990; Zhou, 1999), meaning that the network will model the relative importance of the input data surfaces during the training process without requiring operator input. This characteristic is particularly critical when considering multisource data, as *a priori* knowledge of the level of importance of data layers is not required. A neural network will set weightings to account for a data layer's importance during the training process (Zhou, 1999). These factors make neural networks an appropriate method of classification for this study.

The most commonly used neural network classifier is the multi-layer perceptron (MLP) (Erbek *et al.*, 2004; Liu and Wu 2005). This network has been used previously in remote sensing studies to derive thematic uncertainty measures (Gong *et al.*, 1996; Foody, 2000).

Another network less frequently used in remote sensing than the MLP is the probabilistic neural network (PNN) proposed by Specht (1990). The PNN was tested in this study as it is a non-parametric method of outputting posterior probabilities for every class (Specht, 1990; Tresp *et al.*, 1997; Hart *et al*, 2001), making it particularly suitable for deriving thematic uncertainty measures.

4.4.3 Multi-Layer Perceptron (MLP)

The basic unit of the MLP is the node (Figure 4.2), which mimics a biological neurone. The node sums the inputs and performs a function on the summed input.

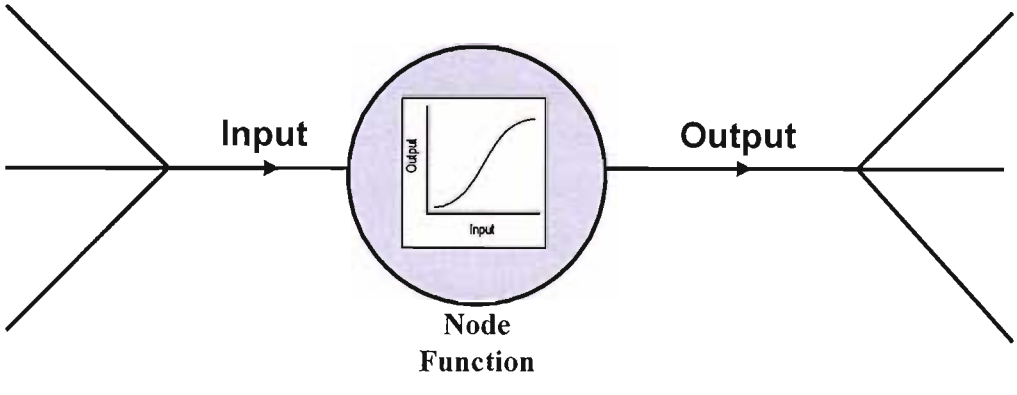


Figure 4.2 Neural network node.

The MLP consists of three types of layers: input, hidden and output (Figure 4.3). The input layer has as many nodes as there are input data layers. There may be one or more hidden layers with the number of layers and nodes specified by the user. The output layer contains as many nodes as there are output classes. Every node in the hidden and output layers is connected to all nodes in the previous layer. As the signal passes between nodes it is modified by weights specific to each node-node connection.

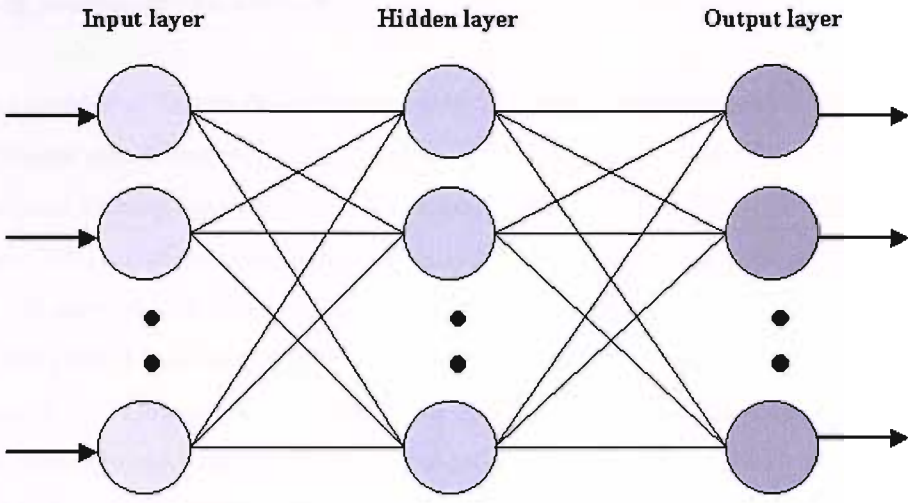


Figure 4.3 Multi Layer Perceptron Neural Network. The connections between the nodes are weighted.

Input signals are passed through the MLP, being modified by the weights associated with the connection between nodes and the functions of each node. The movement of input signals and their modification through the network from input to output is the 'feed-forward' stage of the MLP. The outputs of the MLP are activation levels at each output node. These activation levels may be linked to a biophysical property or land cover class. Training data are entered into the NN and the activation level of each of the output nodes is compared with the input values and an error function is calculated. A learning algorithm is applied that alters the weightings within the network to minimise the error. The whole process is then repeated until a specified number of iterations have taken place, or the error is minimised or reduced below a predetermined level. The alteration of weights through the 'back-propagation' of the error through the network is carried out by the generalised delta rule (Atkinson and Tatnall, 1997; Kanellopoulos and Wilkinson, 1997). This process allows the network to 'learn' the characteristics of the training data set. The number of iterations used in training can affect the accuracy of the final classification (Kavzoglu and Mather, 2003).

As the number of iterations is increased the ability of the network to accurately classify the training data is generally increased, but the network may become less accurate in classifying the main data set (Atkinson and Tatnall, 1997). This effect is known as over-training and results in a loss of the network's ability to classify data it has not seen before (Atkinson and Tatnall, 1997; Kavzoglu and Mather, 2003). The ability of a neural network to accurately classify data that are not used in the training process is known as generalisation (Atkinson and Tatnall, 1997; Foody and Arora, 1997; Kavzoglu and Mather, 2003). This is an important consideration when constructing and training the network.

There are a number of factors that affect the ability of a network to generalise and therefore optimise overall classification accuracy. These include the architecture of the network, the training set and training time (Atkinson and Tatnall, 1997; Foody and Arora, 1997; Kavzoglu and Mather, 2003). The structure of a network is crucial to its classification accuracy. Generally the larger the network, the more accurate it is at classifying the training data (Kavzoglu and Mather, 1999), but it may be less able to generalise. However, a study by Paola and Schowengerdt (1997) found that the number of nodes in a single hidden layer could be varied a great deal with only minimal effects on classification accuracy. This is likely to be data dependent and should be tested with each new dataset used. Increasing the number of input data layers often increases classification accuracy but this is not always the case (Arora and Foody, 1997; Foody and Arora, 1997; Kanellopoulos and Wilkinson, 1997). The training process may be made more complex, as the number of nodes within the network has to be increased to deal with the increased dimensionality (Kanellopoulos and Wilkinson, 1997). If there is correlation between input data layers it is possible that classification accuracy may be reduced compared to using a dataset with no correlation (Kanellopoulos and Wilkinson, 1997; Foody and Arora, 1997). The MLP provides an activation level for every output class of each pixel. In a hard classification the pixel is allocated to the class with the largest activation level. However, the activation levels for all classes may be used to provide additional information for each pixel.

4.4.4 Uncertainty measures from the MLP

MLP activation levels have been used as indicators of class membership on a per-pixel basis (Gong *et al.*, 1996). Gong *et al.* (1996) generated per-pixel measures of thematic uncertainty by normalising activations so that the total output from every pixel summed to one, where a pixel with a large normalised activation was assumed to have a large probability of correct class allocation. However, there are other metrics that may be used as indicators of local thematic uncertainty. Foody (2000) reasoned that MLP activation levels could be used to derive

additional metrics in the confidence of correct pixel allocation in a hard classification. Three main indicators of confidence were suggested; sum of activation levels, maximum activation levels and entropy (Foody 2000). According to the study (Foody, 2000), a pixel was likely to be correctly allocated when the total activation and the maximum activation were approximately one and was more likely to be incorrectly allocated as both these activations tended away from one.

Entropy is a measure of uncertainty and is calculated from the class membership probabilities using the following equation:

$$H = -\sum_{x} p(w) \log_2 p(w) \qquad 4.1$$

where $p(w)$ is the class membership and in the case of the MLP the normalised output activation is assumed to be $p(w)$.

Entropy is maximised when the membership of all classes is equal and minimised when membership is to one class only. The relative entropy (ratio of the observed to the maximum entropy) may be used to assess confidence in a classification. Pixels with small relative entropy are assumed to be associated with one class. Those with large values are associated with multiple classes and a hard classification is not appropriate (Maselli *et al.*, 1994). Maselli *et al.* (1994) did not quantify the error, though Foody (2000) did show a correlation between the entropy of the classification and ground data, indicating that the classifier could identify mixed pixels.

The MLP has the potential to provide measures of thematic uncertainty using a variety of variables (Gong *et al.*, 1996; Foody, 2000). This additional information may be used in change detection to generate probabilities of change, increasing the information content provided.

Maximum activation, sum of activations and entropy provide a single measure of thematic uncertainty for every pixel. Normalised activation provides a measure of thematic uncertainty for every class at every pixel. This makes normalised activation a more suitable measure of thematic uncertainty than the others described above, as the probability of change from any class to any other class can be derived during change detection.

4.4.5 Probabilistic Neural Network (PNN)

4.4.5.1 Bayesian statistics and kernel-based derivation of probability density functions

The purpose of classification is to derive membership of a thematic class from a vector u in feature space corresponding to the sum of the one-dimensional vectors for each waveband of a given pixel. In a Bayesian framework, the method by which this is carried out is by calculating two values for each class, the prior probability and the class conditional probability density function. The prior probability, $P(C_k)$, is the probability that given no other information, a pixel selected at random will belong to class C_k. By reversing this to apply to a classification problem, it may be seen that given no other information a pixel should be allocated the class with the greatest prior probability to maximise the probability of correct classification (Bishop, 1995).

The class conditional distribution, $p(u|\ C_k)$, represents the density function of u, given that a pixel belongs to class C_k. In a classification context the class conditional distribution may be derived from sample data for class C_k.

The proportion of pixels within the image that belong to class C_k and have vector u ($P(C_k,u)$) is:

$$P(C_k,u) = P(C_k).p(u|\ C_k) \qquad 4.2$$

and

$$P(C_k,u) = p(u).P(C_k|u) \qquad 4.3$$

where $p(u)$ is the unconditional density function, the density function for u irrespective of class

$P(C_k|u)$ is the probability that the class is C_k given u and is known as the posterior probability.

Combining 4.2 and 4.3 results in:

$$P(C_k|u) = \frac{p(u|C_k).P(C_k)}{p(u)} \qquad 4.4$$

where $p(u)$ normalises the posterior probability.

The posterior probability is derived from a density function known as the probability density function (PDF), which is represented by the right-hand side of Equation 4.4. The PDF may be derived from training data allowing classification to be carried out on the basis of posterior probabilities.

However, a parametric Bayesian approach to classification may make assumptions about the data used that are not valid. Remotely sensed data may have skewed or multi-modal distributions, rendering a parametric approach to classification invalid. A non-parametric approach to deriving the PDF from sample data was suggested by Parzen (1962). This involves placing simple functions or kernels at locations in feature space derived from samples of the population. The integral of these kernels may be used to provide an approximation of the PDF. As the number of samples is increased, the approximation of the PDFs asymptotically approaches the population density function (Specht, 1990). This approach allows a PDF to be derived with no previous knowledge of the data distribution.

4.4.5.2 Probabilistic Neural Networks

One limitation of the method proposed by Parzen (1962) is that the entire training set must be stored and used during classification and the amount of computation required is proportional to the size of the training set. At the time of Parzen's (1962) paper, the processing power required to calculate PDFs using kernels was not available. As computational power increased, this method became more practical and Specht (1990) proposed a neural network that used kernel functions to calculate PDFs. This probabilistic neural network (PNN) used Gaussian radial functions rather than the sigmoid functions commonly used by networks such as the MLP.

The PNN is a feedforward network that has three layers in the same layout as a MLP with a single hidden layer (Figure 4.3): input, pattern and output. The input layer contains as many nodes as there are input data layers. The pattern layer corresponds to the hidden layer in the MLP and has as many nodes as there are training pixels. Each node models a kernel based on the point represented in feature space by the training pixel. The output layer contains as many layers as there are classes. Each node in the radial layer is only connected with the class output node associated with the training data and sums the inputs from the pattern layer.

When the allocation stage of classification is carried out, the probability of membership to each of the radial nodes is calculated and these are summed for each class and normalised to give a

posterior probability of membership to each class. The pixel is then allocated to the class it has the largest posterior probability of membership to.

In the simplest form of PNN as described above, it is assumed that pixels are randomly selected from the complete population. However, if the proportions of pixels used for each class are not equivalent to the prior probability then weights may be applied to compensate for the differences between the training sample sizes and the prior probabilities. This is of particular use when some of the classes used have relatively small prior probabilities. If random sampling were used, the number of training samples for classes with small prior probabilities would be small, potentially resulting in a sample that could not model the PDF accurately. To reduce this effect, the number of training pixels may be increased and a weighting to account for prior probabilities applied to the output in order to provide more accurate posterior probabilities. One method is to use an equal number of training samples for each class and estimate the prior probabilities from ground sampling or image interpretation.

The width of the kernel distribution is determined by the operator and is known as the smoothing function, h. When the smoothing function is too large, the estimated PDF is over smoothed, resulting in an inaccurate classification (Bishop, 1995). As $h \to 0$ the PDFs will approach an exact representation of the density, for an infinite sample size. However, for a finite sample, as $h \to 0$ the PDF will approach a set of delta functions representing each training sample, resulting in a noisy representation of the PDF. When $h = 0$ and assuming the training samples used for different classes represent different points in feature space, the classification accuracy of the training data will be close to 100%. However, any point in feature space not represented in the training data will not be classified, resulting in an inability of the network to generalise. This characteristic of PNNs means that care must be taken when determining the smoothing function to be used. The training error may give an inaccurate indication of the ability of the PNN to correctly classify non-training data. For this reason it is essential that the correct smoothing function should be determined by testing the PNNs with separate data from training data. The effect of h on the PDF may be seen in Figure 4.4.

PNNs have advantages over networks that are trained iteratively, as training only requires generation of kernels for each of the training pixels. The only variable that needs to be determined is the smoothing function and this may be determined by experimentation. However, each training point is represented by a node in the radial layer and so the allocation process can be very intensive computationally, especially if large training samples are used.

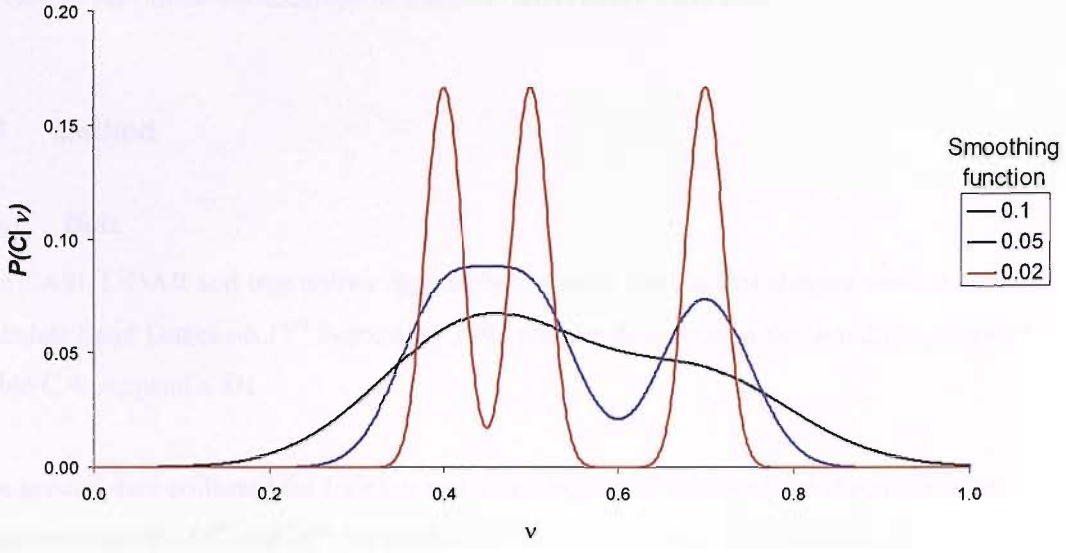


Figure 4.4 Example of the kernel approach to density estimation for class *C* using vector v. One-dimensional Gaussian kernels. Training samples at 0.4, 0.5 and 0.7.

The generation of the PDFs for each class by PNNs mean that the outputs may be interpreted directly as posterior probabilities and, therefore, may be used to derive per-pixel uncertainties.

Though there are few remote sensing studies that have used PNNs, they have been used for classification of magnetometry data for the detection of buried unexploded ordnance (Hart *et al.*, 2001), texture classification (Raghu and Yegnanarayana, 1998), cloud classification (Tian and Azimi-Sadjadi, 2001; Wang *et al.*, 2004), land use classification (Ashish *et al.*, 2004) and detecting classes that are not represented in training data during land cover classification (Augusteijn and Folkert, 2002).

4.5 Neural networks for deriving thematic uncertainty

This study will use neural networks to derive per-pixel thematic uncertainty measures. When setting up these networks there are various network variables that are defined during the training stage that can alter the outputs. These variables influence the thematic accuracy and the thematic uncertainty measures derived. In the case of the MLP these network variables are the number of nodes in the hidden layer, the learning rate, momentum rate and number of iterations. In the case of the PNN the only network variable is the smoothing function. Though previous studies have derived heuristics for maximising thematic accuracy (Arora and Foody, 1997; Foody and Arora, 1997; Kavzoglu and Mather, 2003), heuristics have not been derived for maximising the accuracy of thematic uncertainty measures. In this chapter the derivation of

thematic uncertainty measures using neural networks will be examined and heuristics will be derived to maximise the accuracy of thematic uncertainty measures.

4.6 Method

4.6.1 Data

The CASI, LiDAR and true colour digital photography used in this chapter were acquired over Ainsdale Sand Dunes on 11th September 2002 and are described in Section 2.6.1 (Appendix C; Table C.4; Appendix D).

The ground data collected for training and assessing the accuracy of the classifiers were collected between 16th and 20th September 2002 and are described in Section 2.6.2.

4.6.2 Training the neural networks

1000 training pixels per class were used to train the networks. For each class equal numbers of pixels were selected at random from each training polygon.

Trajan 6.0 (Trajan, 2001) software was used to generate the MLPs. MLP networks were generated with a range of nodes between 5 and 40 in a single hidden layer. Each network was trained using between 250, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500 and 4000 iterations. To reduce the total number of networks tested, a fixed learning rate and momentum of 0.1 and 0.3 respectively were used for all architectures. These values had been used in previous studies using these data (Brown *et al.*, 2003a) and did not result in large oscillations in error during training and networks did not stick in local error minima during the training process.

Due to a bug in Trajan 6.0 when generating PNNs, Trajan 4.0 was used to generate these networks. The nodes in the pattern layer of the PNNs generated modelled Gaussian distributions with a σ value equal to the smoothing function. PNN networks were generated with the following range of smoothing functions: 0.0025, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, and 0.2. Prior probabilities for use in the PNN were estimated by photo interpretation of 786 points using a geographically stratified random sampling strategy, with the same units as the accuracy assessment data (60 m; Section 2.6.2), but separate randomly selected points within each stratum (Table 4.1).

Table 4.1 Estimated prior probabilities used in PNN training.

Water	Sand	Marram	Grass	Reeds	Creep	Buckthorn	Woodland
0.06870	0.08015	0.03944	0.37405	0.02290	0.09542	0.01018	0.30916

4.6.3 Thematic uncertainty prediction

The ability of the networks to predict thematic uncertainty was tested in two ways:

1. Class independent thematic uncertainty.
2. Class specific thematic uncertainty.

4.6.3.1 Class independent thematic uncertainty

MLP and PNNs produce an activation from each output node that has the potential to provide an indication of the thematic uncertainty for every class on a per-pixel basis. The link between every output activation and thematic uncertainty was tested using all output activations. The activation for every output of every pixel was averaged within the following bins $0-0.0\dot{9}$, $0.1-0.1\dot{9}$, $0.2-0.2\dot{9}$, $0.3-0.3\dot{9}$, $0.4-0.4\dot{9}$, $0.5-0.5\dot{9}$, $0.6-0.6\dot{9}$, $0.7-0.7\dot{9}$, $0.8-0.8\dot{9}$, $0.9-1.0$. The mean of the activation and the proportion of times that any of the activations within the bin was the correct class were calculated for each bin. An RMSE value was calculated from mean activations and proportion of correct classes for all networks tested and an *F*-test was carried out assuming that activation was equal to proportion of times a class was correct. The RMSE measure derived provides an indication of how accurately the network predicts thematic uncertainty, independent of class. This error statistic was defined as the class independent thematic uncertainty RMSE.

4.6.3.2 Class specific thematic uncertainty

It was essential for the change detection process that a network was able to predict thematic uncertainty for all classes. If this did not occur, the output from change detection would not be an accurate representation of the probability of change for classes that were not accurately classified. To provide a measure of whether the networks accurately predicted thematic uncertainty for every class, the following procedure was carried out. For a specific class, the proportion of times that pixel allocated to that class was correct (the user's accuracy) and the mean activation for all pixels allocated to that class were calculated. A RMSE value was calculated from the difference between the user's accuracy, and the mean activation for all classes. An *F*-test was carried out assuming that the user's accuracy should equal the mean

activation for all classes. The RMSE measure derived provided an indication of how accurately the network predicted thematic uncertainty, dependent on class and was defined as the class specific thematic uncertainty RMSE.

4.6.3.3 Overall entropy

To determine how network activations change with network variables, a global measure of entropy was used (Equation 4.1). Entropy is minimised when activations tend towards 0 and 1. Entropy provided an indication of the confidence of the network in the classification produced. This network confidence was not an indication of the probability of correct classification, and did not necessarily provide a measure of uncertainty.

4.6.4 Effect of neural network setup and training variables

To determine whether there were heuristics that could be used to select the network that classified and predicted thematic uncertainty most accurately the correlations between the variables used when setting up and training MLPs and PNNs and thematic accuracy and thematic uncertainty prediction were tested. The correlation was tested using linear, log-linear and second order polynomial relationships in order to determine whether there were relationships between network variables and thematic accuracy or thematic uncertainty. These regression models were used as they could provide an indication of simple relationships for deriving heuristics to maximise thematic accuracy and the accuracy of thematic uncertainty measures. The correlation with the largest R^2 was assumed to be most representative of the relationship.

4.7 Results

4.7.1 MLP

All MLP results are in Appendix G.

4.7.1.1 Global thematic accuracy

Of the MLPs tested, two had the maximum overall accuracy (P_o). Both networks had 25 nodes in the hidden layer, with one trained for 2000, the other 2250 iterations (Figure 4.5).

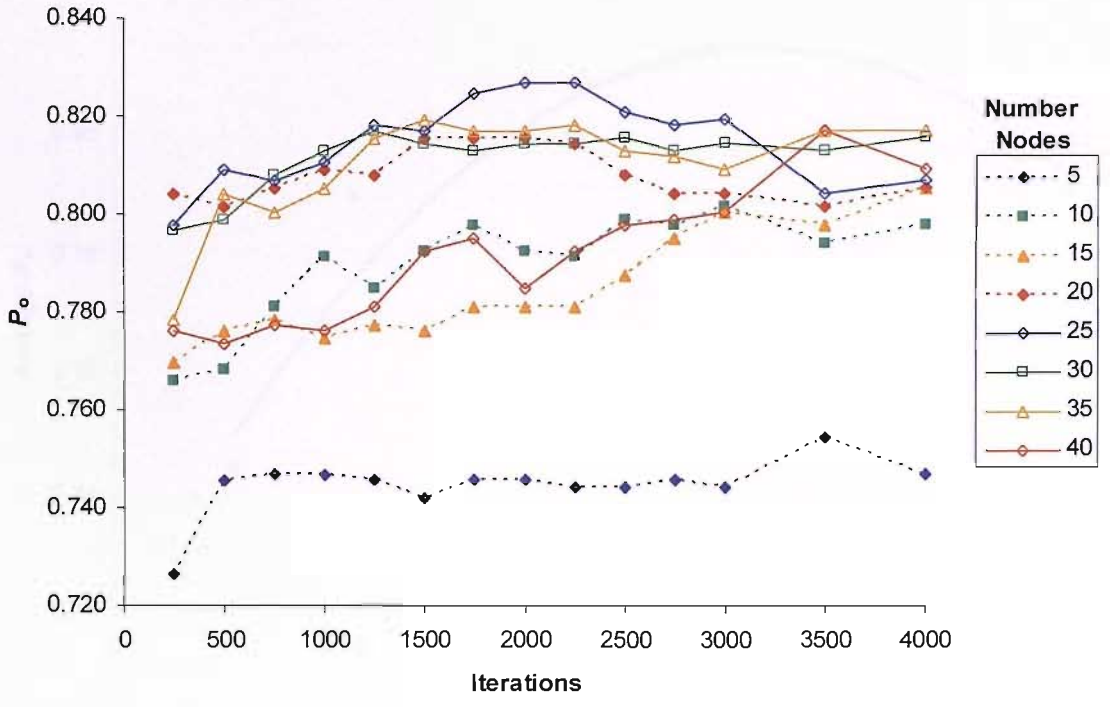


Figure 4.5 MLP P_o as a function of iterations and architecture.

When the P_o was averaged for all networks with a given number of nodes, it may be seen that accuracy was maximised when the number of nodes in the hidden layer was 25 (Figure 4.6). Though this result showed that beyond a certain number of nodes the overall accuracy decreased, the decrease in accuracy beyond this point was exaggerated using average P_o. At 40 nodes it appeared that there was a large reduction in accuracy using the averaged P_o. Using a second order polynomial regression this relationship was found to be significant (R^2=0.866, F=23.7, p=0.003).

A general increase in P_o with an increase in the number of iterations may be seen when P_o was averaged for all networks at a given iteration (Figure 4.7). When tested using a log linear regression a significant correlation was found (R^2=0.949, F=243.8, p<0.001).

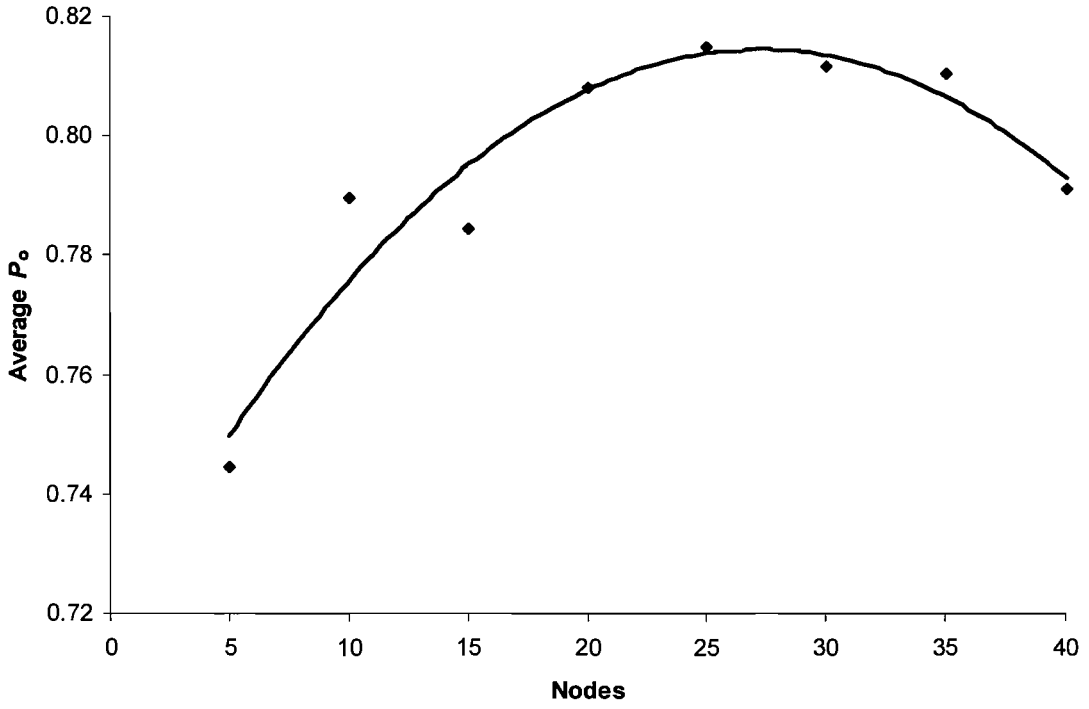


Figure 4.6 Average P_o as a function of number of nodes in MLP hidden layer. Second order polynomial regression line.

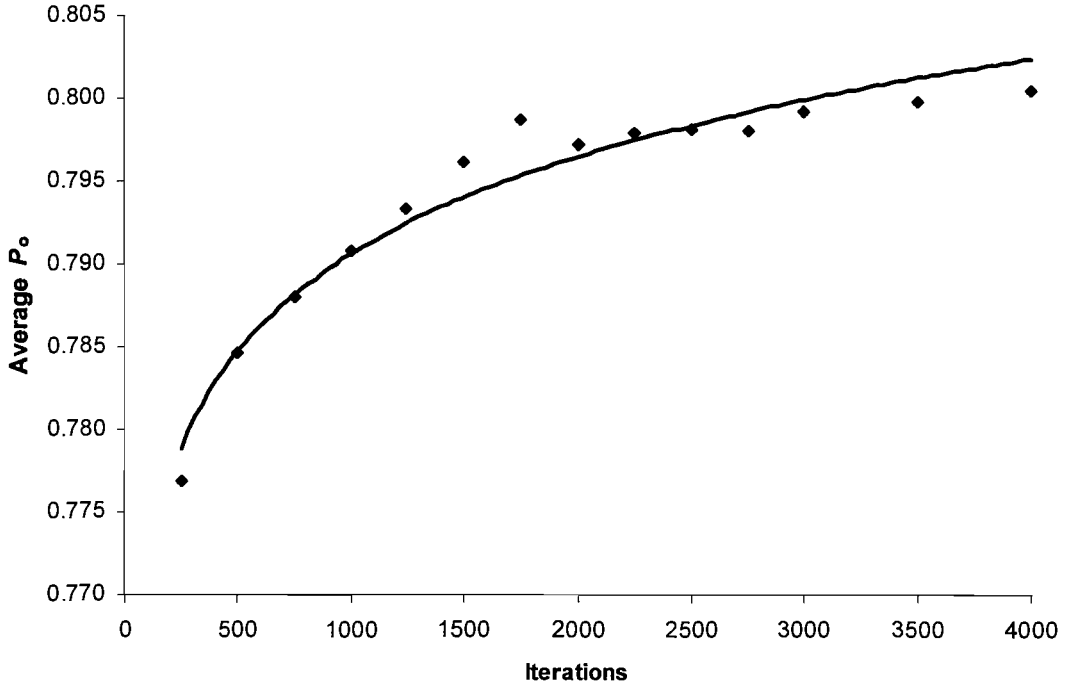


Figure 4.7 Averaged P_o as a function of iteration for all networks tested. Log linear regression line.

4.7.1.2 Thematic uncertainty

The output activation levels were tested for suitability as uncertainty measures, using the assumption that the output activation equalled probability of correct allocation. A linear regression was carried out to determine the correlation between the predicted thematic uncertainty in the form of output activation and the actual thematic uncertainty in the form of the proportion of correct allocation. For the MLP the activation values for each pixel were normalised to sum to one.

For all MLPs tested, the output activation and class independent thematic uncertainty were correlated and the relationship was significant (at 95% confidence) (Appendix G; Table G.4 ; Table G.5). The MLP with 10 nodes in the hidden layer and trained for 1000 iterations had the smallest class independent thematic uncertainty RMSE of all the MLPs tested (RMSE=0.059, R^2=0.982, F=491.4, p<0.001) (Figure 4.8). This indicated that this MLP predicted thematic uncertainty most accurately. The most accurate MLPs (25 nodes, 2000 iterations and 25 nodes, 2250 iterations) had class independent thematic uncertainty RMSE values 36% larger than the value of the smallest RMSE (RMSE=0.092, R^2=0.942, F=147.9, p <0.001) and (RMSE=0.095, R^2=0.943, F=150.1, p<0.001) respectively (Figure 4.8).

The architecture that resulted in the most accurate estimation of thematic uncertainty was ten nodes in the hidden layer when the average class independent thematic uncertainty RMSE was considered (Figure 4.9). There was an overall increase in class independent thematic uncertainty RMSE above and below this value (Figure 4.9). However, when this relationship was tested using all regression relationships no significant correlation was found (at 95% confidence) (Appendix G; Table G.7).

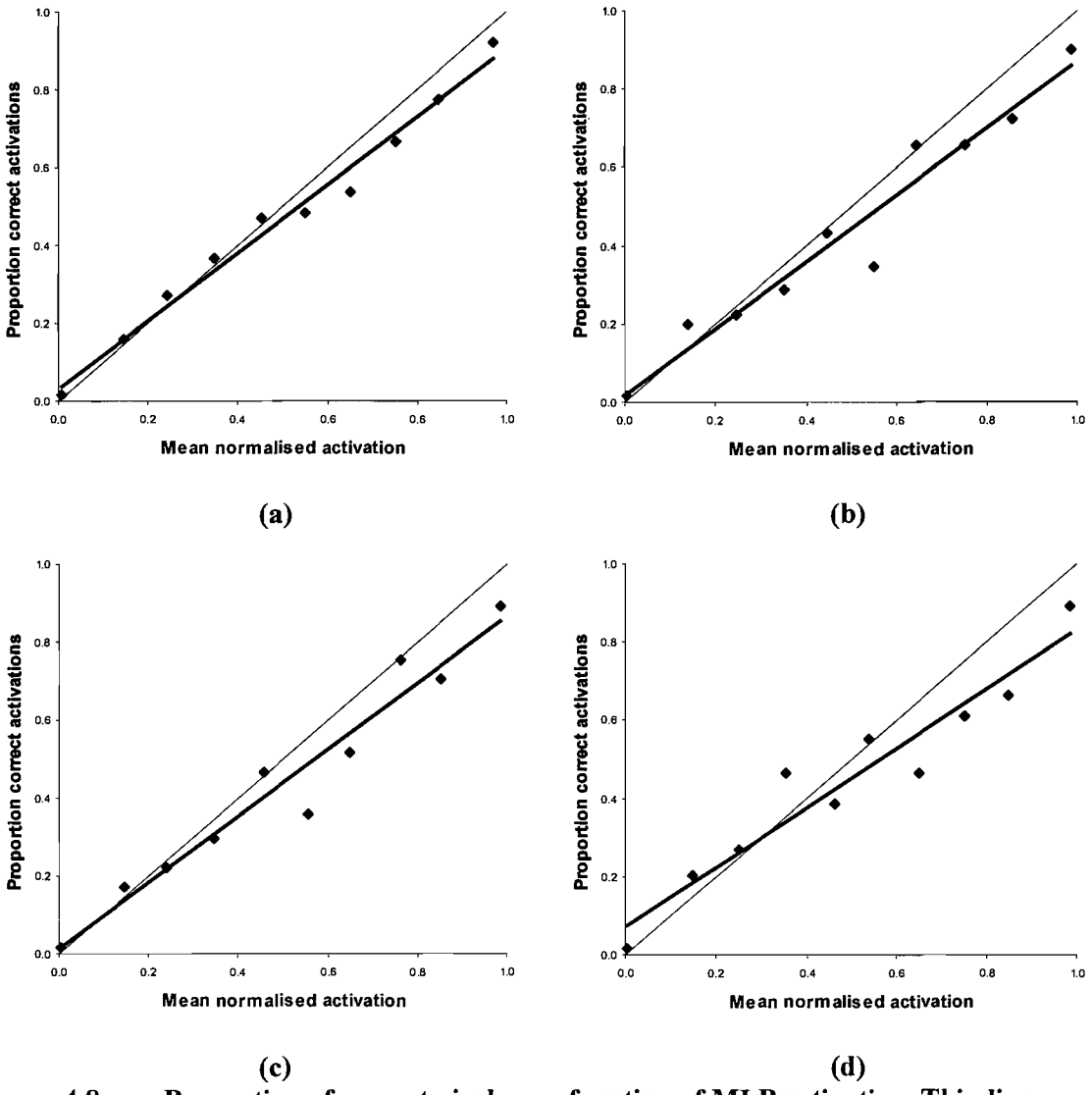


Figure 4.8 Proportion of correct pixels as a function of MLP activation. Thin line represents assumption of one to one relationship. Thick line represents linear regression.
a) 10 nodes, 1000 iterations.
b) 25 nodes, 2000 iterations.
c) 25 nodes, 2250 iterations.
d) 30 nodes, 3500 iterations.

Class independent thematic uncertainty RMSE and the number of iterations were found to be correlated using second order polynomial regression (R^2=0.324, F=4.12, p=0.046) (Figure 4.10). Above 1000 to 1500 iterations, class independent thematic uncertainty RMSE tended to increase with number of iterations. When the number of iterations was less than this number, there was a general trend for class independent thematic uncertainty RMSE to increase with a decrease in the number of iterations.

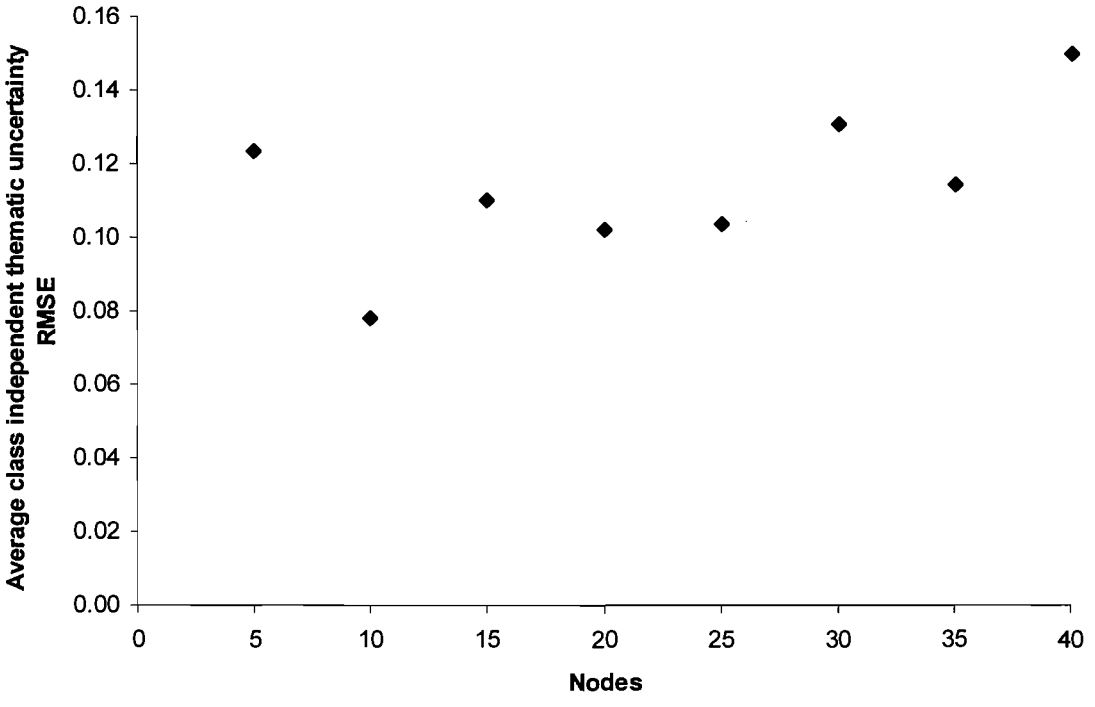


Figure 4.9 Average class independent thematic uncertainty RMSE as a function of nodes in hidden layer.

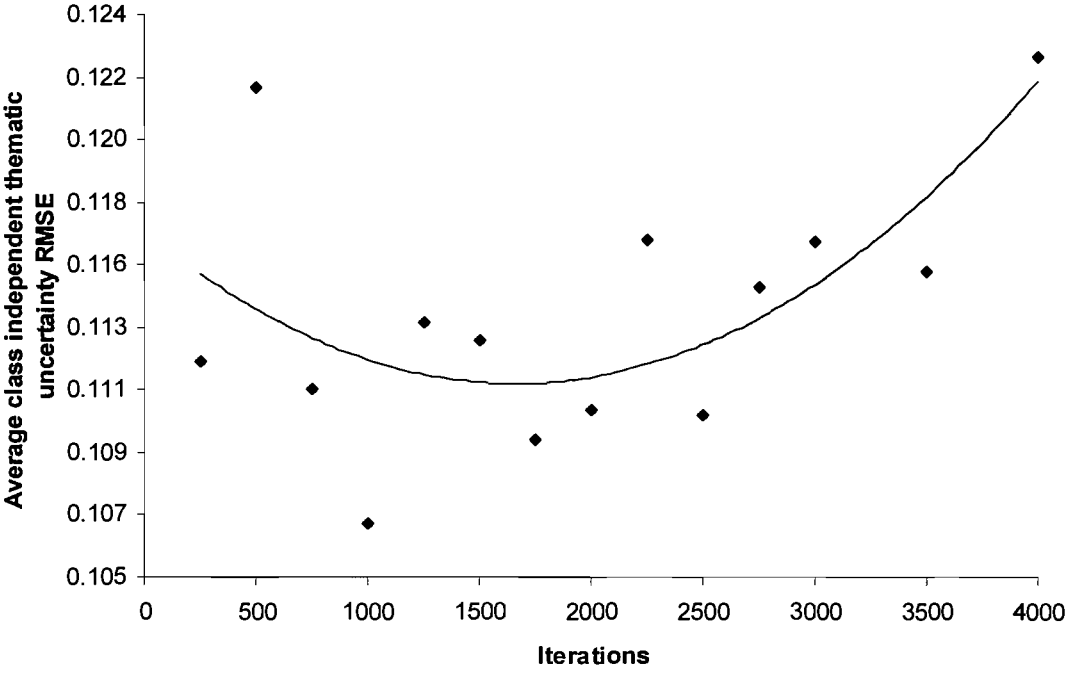


Figure 4.10 Average class independent thematic uncertainty RMSE as a function of iterations. Second order polynomial regression line.

To determine whether there was a correlation between the ability of MLPs to accurately classify data and predict class independent thematic uncertainty, P_o and RMSE values for all the

networks were compared (Figure 4.11). Using linear regression and an F-test it was determined that there was no significant correlation between these variables (at 95% confidence) (Appendix G; Table G.8).

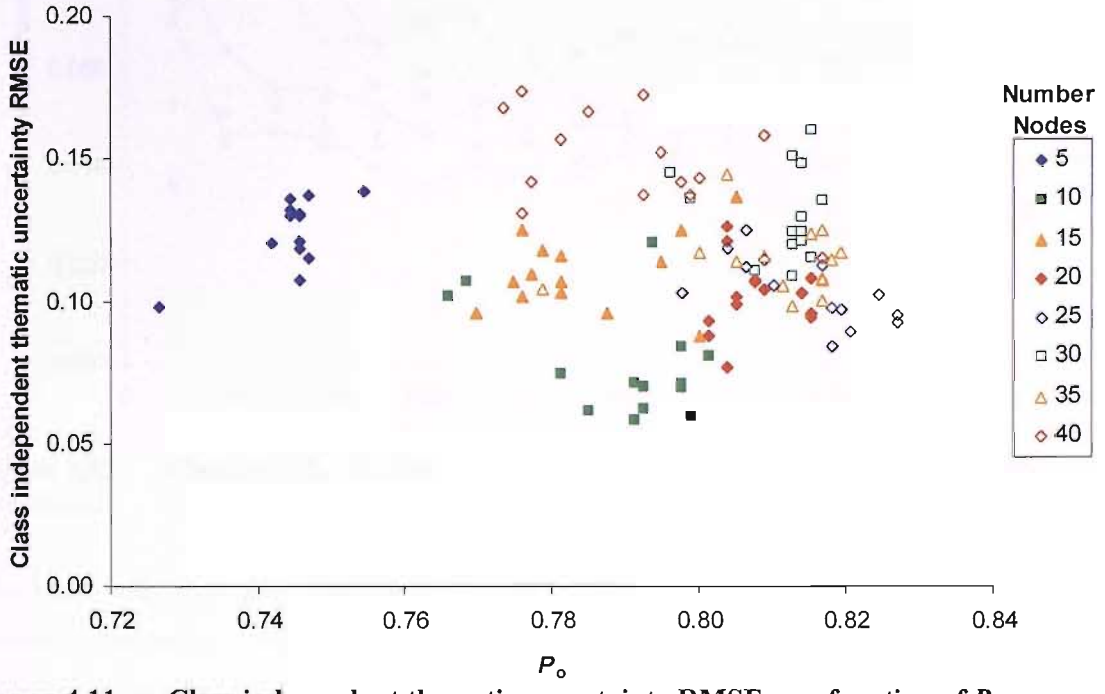


Figure 4.11 Class independent thematic uncertainty RMSE as a function of P_o.

Though the class independent thematic uncertainty RMSE provided a measure of how accurately MLP activations represented thematic uncertainty, it did not provide a measure of how accurately activations represented uncertainty for all of the classes. The class specific thematic uncertainty RMSE provided an indication of the accuracy of activations for representing thematic uncertainty on a per-class basis.

Using linear regression it was found that there was significant correlation (at 95% confidence) between predicted and actual class specific thematic uncertainty for all MLPs (Appendix G). The smallest class specific thematic uncertainty RMSE was for the MLP with 30 nodes in the hidden layer and trained for 3500 iterations (RMSE=0.025, R^2=0.985, F=459.7, p<0.001) (Figure 4.12).

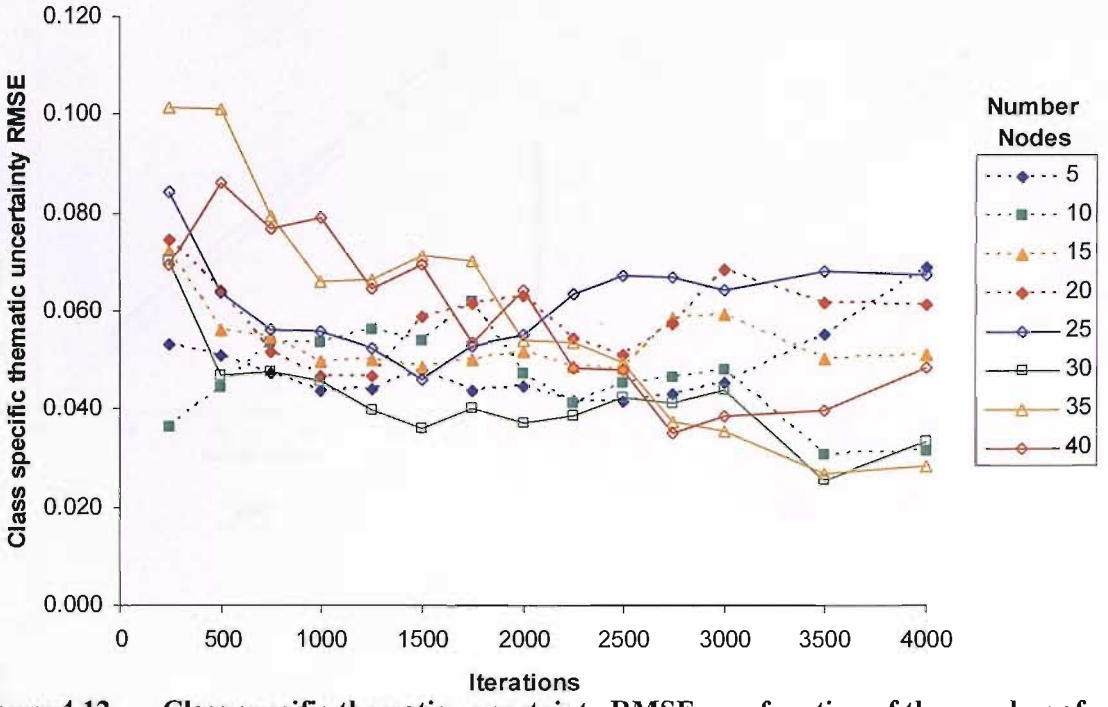


Figure 4.12 Class specific thematic uncertainty RMSE as a function of the number of iterations.

The MLPs with the largest overall accuracy (25 nodes, 2000 iterations and 25 nodes, 2250 iterations) had a class independent thematic uncertainty RMSE values 55% and 61% larger than the value of the smallest class independent thematic uncertainty RMSE respectively (RMSE=0.055, R^2=0.960, F=167.6, p<0.001) and (RMSE=0.064, R^2=0.946, F=124.8, p<0.001) (Figure 4.13).

The MLP that represented class independent thematic uncertainty most accurately (10 nodes, 1000 iterations) had class specific thematic uncertainty RMSE values 54% larger (RMSE=0.054, R^2=0.982, F=392.9, p<0.001) than the value of the smallest class specific thematic uncertainty RMSE.

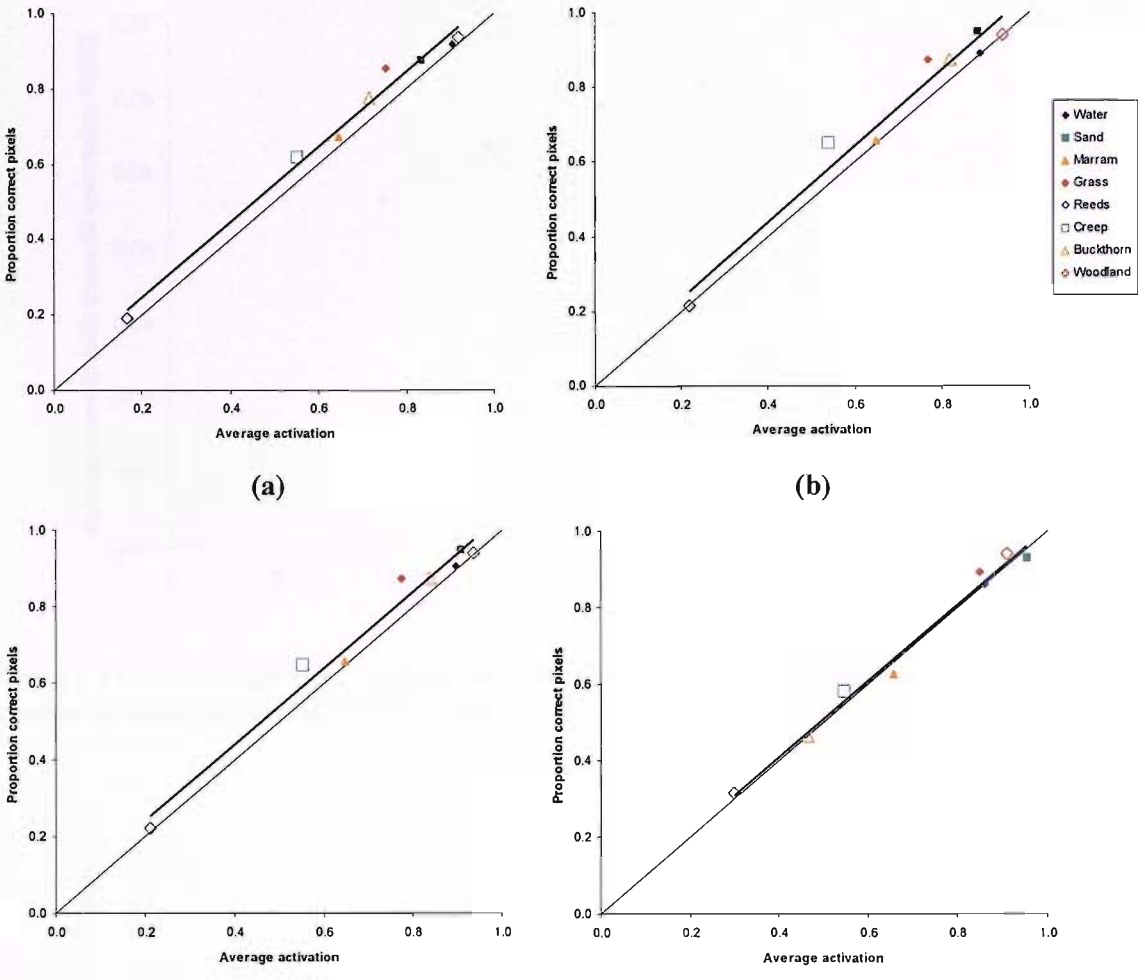


Figure 4.13 Proportion of correct pixels per class as a function of MLP activation. Thin line represents assumption of one to one relationship. Thick line represents linear regression.
a) 10 nodes, 1000 iterations.
b) 25 nodes, 2000 iterations.
c) 25 nodes, 2250 iterations.
d) 30 nodes, 3500 iterations.

A significant relationship between class specific thematic uncertainty error and the number of nodes in the hidden layer (at 95% confidence) was not found (Figure 4.14). From Figure 4.12 it may be seen that for most MLP architectures the trend was for class specific thematic uncertainty RMSE to reduce as the number of iterations was increased. A log-linear regression of averaged class specific thematic uncertainty RMSE values against iterations showed a significant correlation (R^2=0.922, F=155.5, p<0.001) (Figure 4.15).

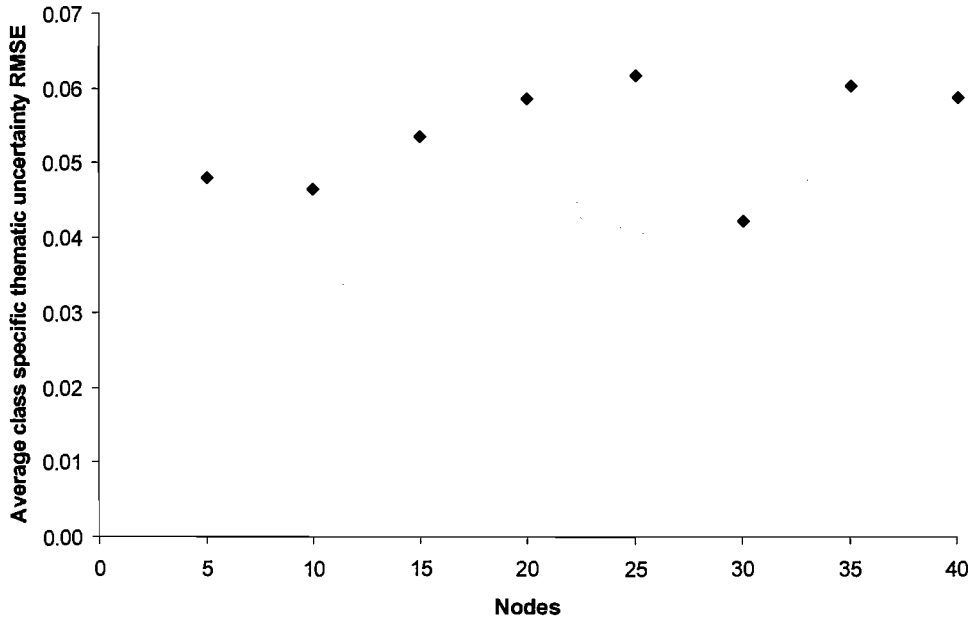


Figure 4.14 Average class specific thematic uncertainty RMSE as a function of the number of nodes in the MLP hidden layer.

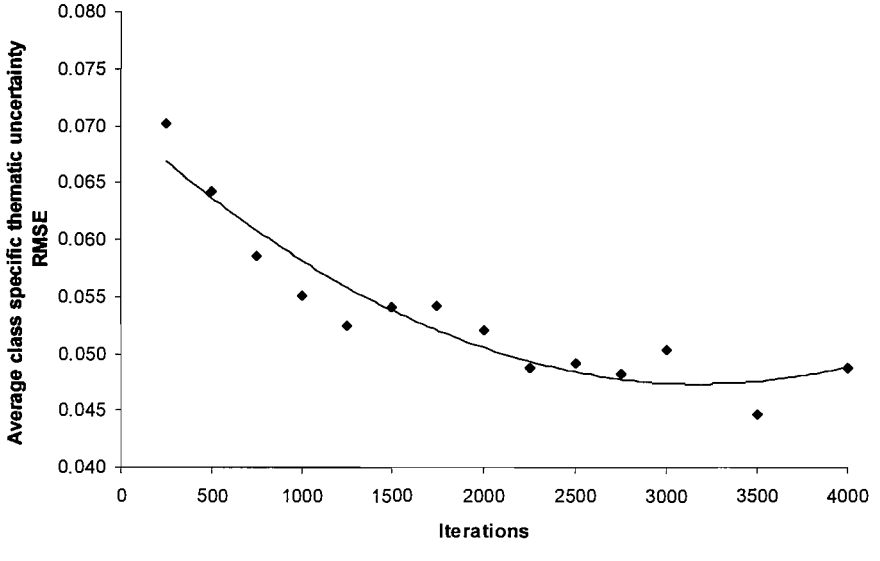


Figure 4.15 Average class specific thematic uncertainty average RMSE as a function of the number of iterations. Log-linear regression line.

When the relationship between entropy and P_o was tested using linear regression it was found that they were negatively correlated and the relationship was significant (R^2=0.579, F=153.9, p<0.001) (Figure 4.16). Therefore as P_o increased, activations tended towards 0 and 1.

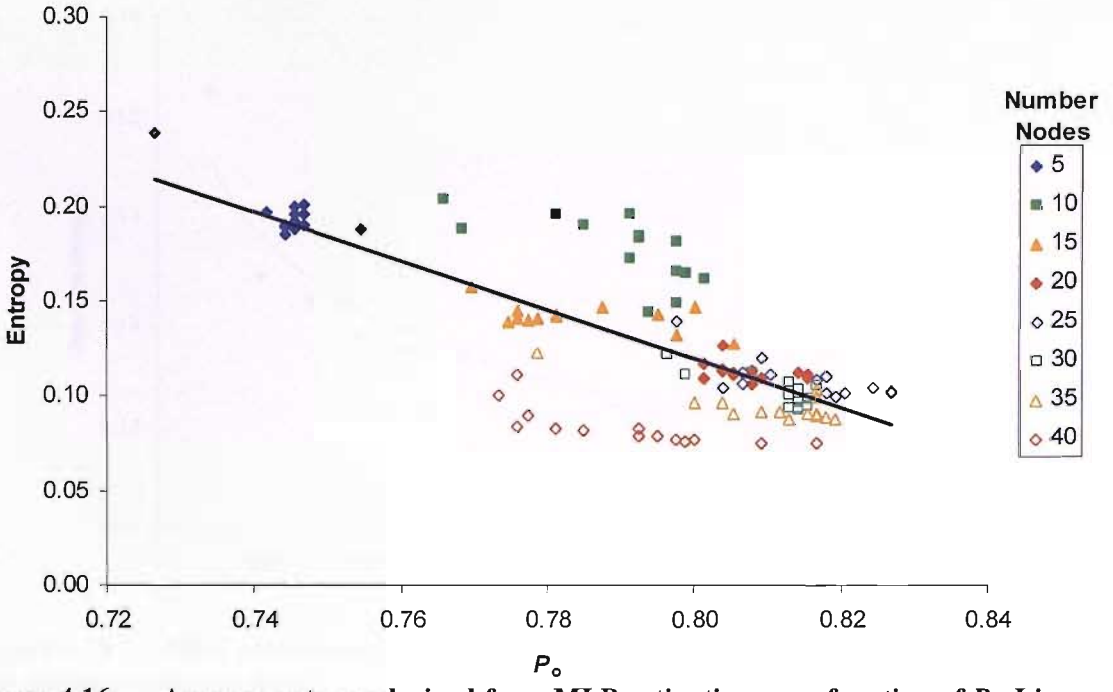


Figure 4.16 Average entropy derived from MLP activations as a function of P_o. Linear regression line.

There was a significant correlation between entropy and number of nodes assuming a log-linear relationship (R^2=0.960, F=168.3, p<0.001) (Figure 4.17) and entropy and number of iterations assuming a log-linear relationship (R^2=0.910, DF=12, F=91.0, p<0.001) (Figure 4.18). MLP entropy generally decreased with an increase in the number of nodes and the number of iterations (Figure 4.17; Figure 4.18).

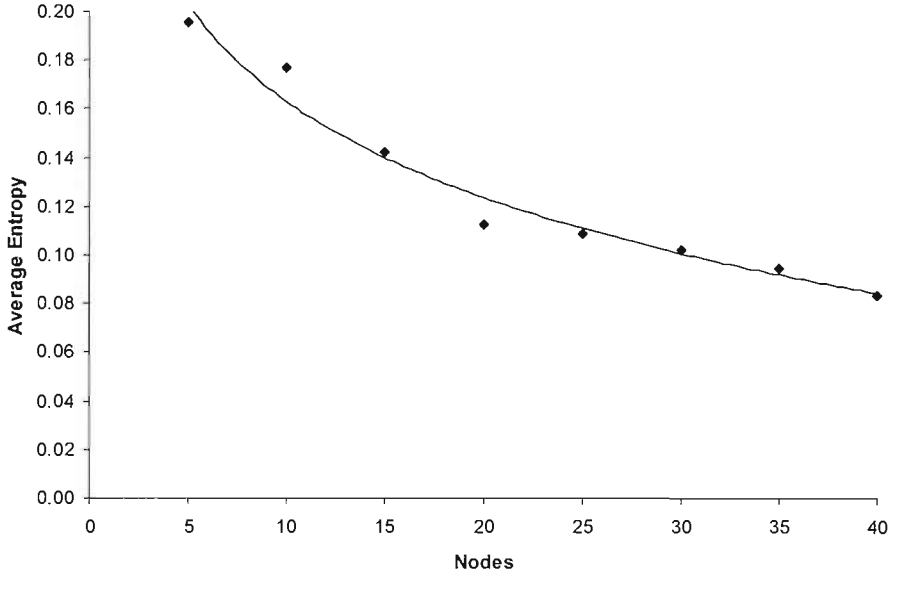


Figure 4.17 Average entropy as function of the number of nodes. Log-linear regression line.

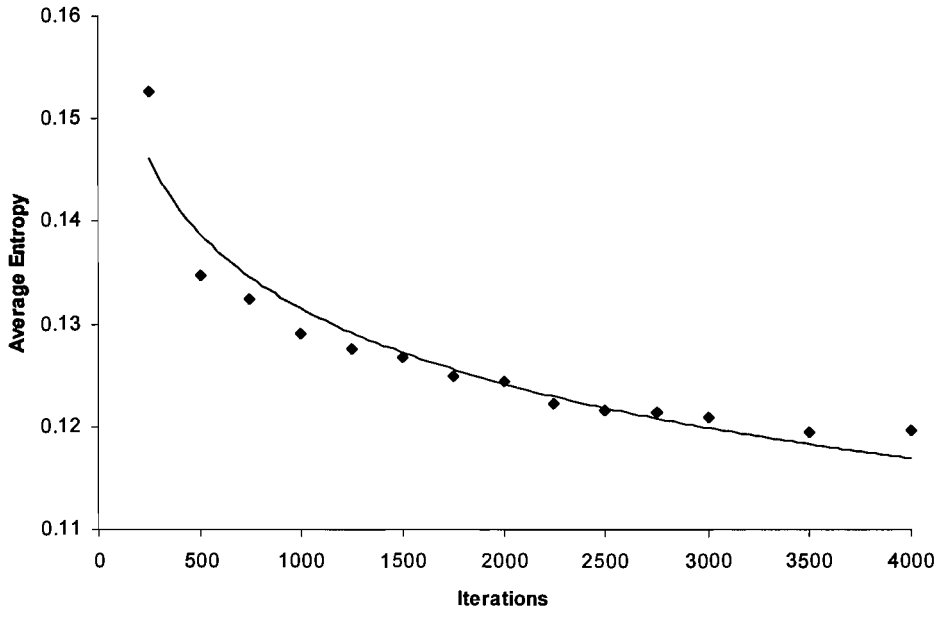


Figure 4.18 MLP activation average entropy as function of the number of iterations. Log-linear regression line.

When the P_o, class independent thematic uncertainty RMSE and class specific thematic uncertainty RMSE were tested for correlation using linear regression there were no significant relationships (at 95% confidence) between any of the variables (Appendix G; Table G.8).

It was not possible to determine the most accurate MLP, as accuracy depended on the measure being used in terms of overall accuracy, class independent thematic uncertainty RMSE and class specific thematic uncertainty RMSE (Table 4.2). Different MLP networks were the most accurate dependent on the specific measure used to define accuracy.

Table 4.2 Most accurate MLP networks dependent on variable used to define accuracy.

Nodes	Iterations	P_o	Class independent thematic uncertainty RMSE	Class specific thematic uncertainty RMSE
10	1000	0.791	0.059	0.054
25	2000	0.827	0.092	0.055
25	2250	0.827	0.095	0.064
30	3500	0.813	0.109	0.025

4.7.2 PNN

All PNN results are in Appendix H.

4.7.2.1 Global thematic accuracy

The PNN with a smoothing function of 0.02 was the most accurate (Figure 4.19). Below this value, a decrease in smoothing function resulted in a large drop in P_o, with P_o less than 0.5 for a smoothing function of 0.0025 (Figure 4.19). Above a smoothing function of 0.02 an increase in smoothing function resulted in a decrease in P_o (Figure 4.19). From this it may be seen that P_o of the PNNs tested was dependent on the smoothing function.

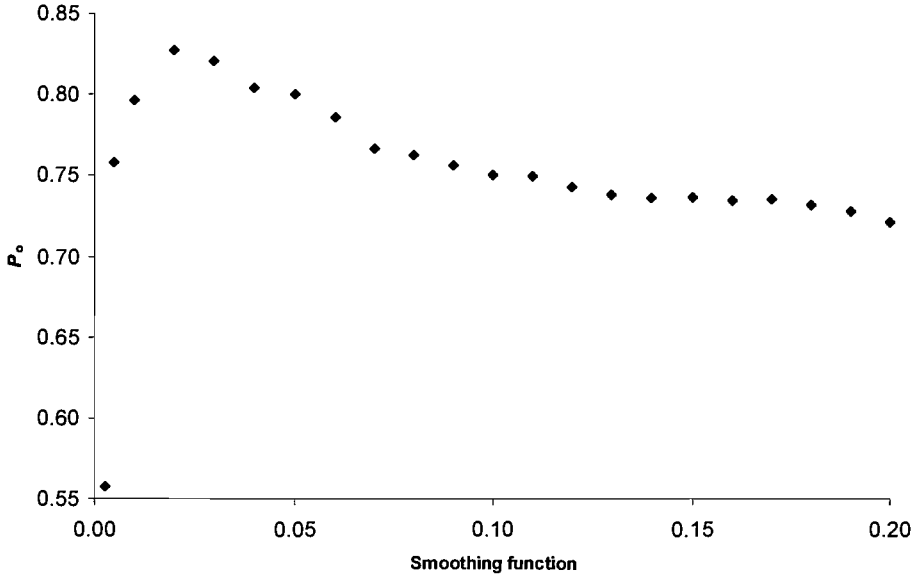


Figure 4.19 PNN P_o as a function of smoothing function.

4.7.2.2 Thematic uncertainty

For all PNNs tested, apart from the one with a smoothing function of 0.0025 (RMSE=0.408, F=3.03, p=0.120), the relationship between output activation and proportion of correct pixels was significant using an F-test (Appendix H; Table H.2).

The PNN smoothing function affected class independent thematic uncertainty RMSE. The smallest class independent thematic uncertainty RMSE occurred when the smoothing function was 0.11 (Figure 4.20; Figure 4.21). For values of the smoothing function above and below 0.11 PNNs were less able to predict thematic uncertainty. At very small values of the smoothing

function the activation predicted the thematic uncertainty less well. With a smoothing function of 0.005 the RMSE was 0.215 (F= 9.71, p<0.05) (Figure 4.20).

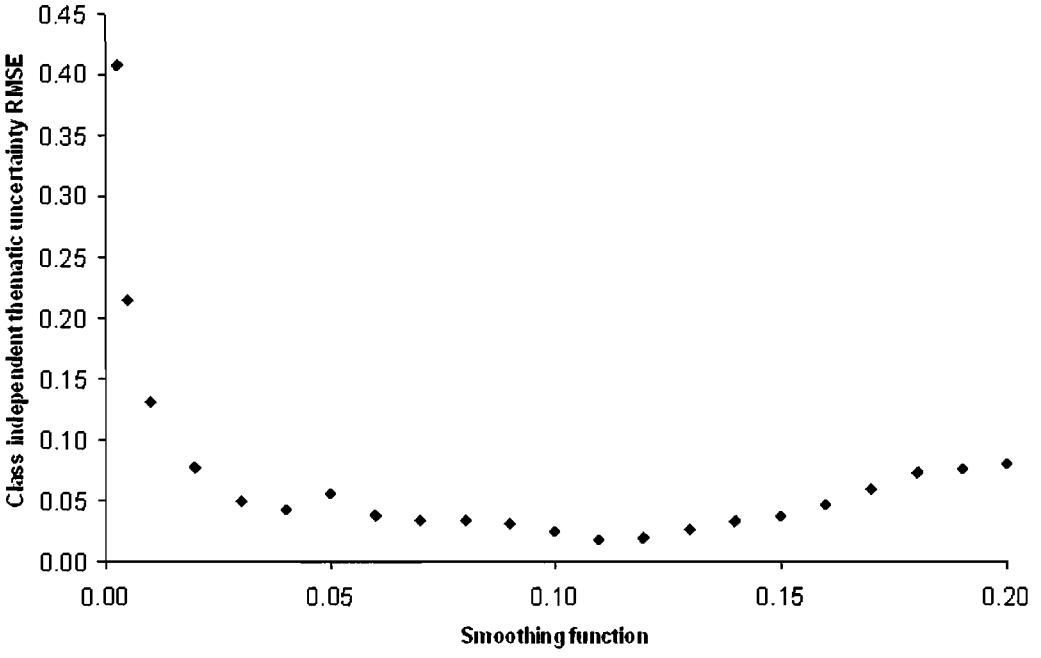
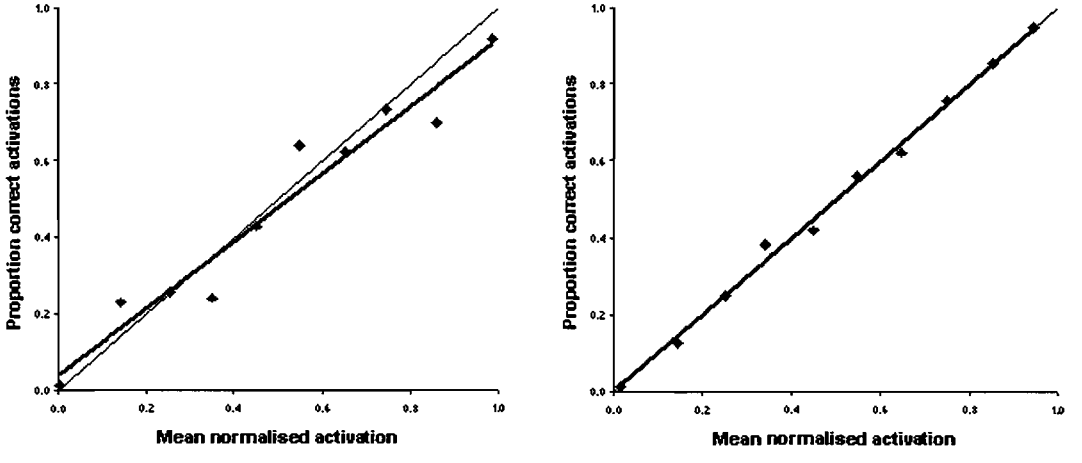


Figure 4.20 PNN class independent thematic uncertainty RMSE as a function of smoothing function.



(a) (b)

Figure 4.21 Proportion of correct pixels as a function of PNN activation. Thin line represents assumption of one to one relationship. Thick line represents linear regression.
a) 0.005 smoothing function.
b) 0.11 smoothing function.

As the PNN smoothing function was increased, entropy increased (Figure 4.22). This was expected, as when the smoothing function was increased there would be greater overlap

between the PDF of classes (Figure 4.4). This would result in activation levels tending away from 0 and 1.

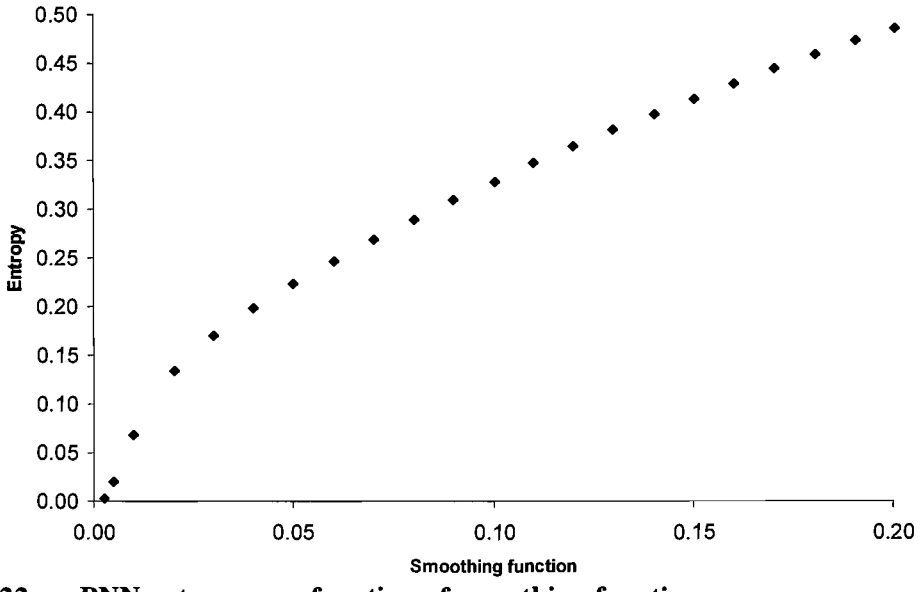


Figure 4.22 PNN entropy as a function of smoothing function.

4.7.2.3 Per-class accuracy

The per-class accuracy was dependent on the smoothing function and the per-class metric used (Figure 4.23; Figure 4.24; Table 4.3; Table 4.4).

Producer's accuracy (Appendix B) was maximised for the Sand, Marram, Reeds, Creep and Buckthorn classes with smoothing functions between 0.005 and 0.02. For all these classes apart from Sand, the Producer's accuracy fell to zero as the smoothing function was increased beyond 0.02. For the Woodland class, Producer's accuracy was maximised for a smoothing function of 0.03. The Grass Producer's accuracy was maximised at a smoothing function of 0.11 and the Water class had 100% Producer's accuracy for all smoothing functions tested apart from 0.005 and 0.01. The Water, Woodland, Sand and Grass classes were less affected by smoothing than the other classes, with only small changes in Producer's accuracy with changes in smoothing function (Figure 4.23). The User's accuracy (Appendix B) was maximised for Sand, Grass and Creep with a smoothing function of 0.0025 (Figure 4.24).

These results show that class accuracy was related to smoothing function, but the smoothing function that resulted in the most accurate class measures varied for each class.

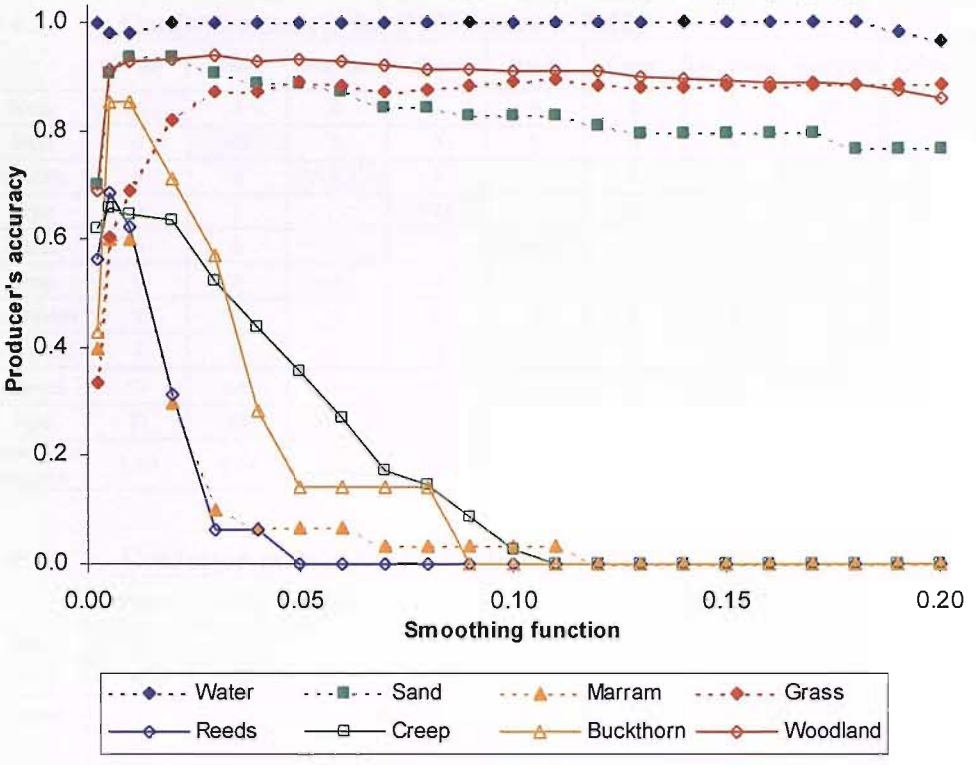


Figure 4.23 **PNN Producer's accuracy as a function of smoothing function.**

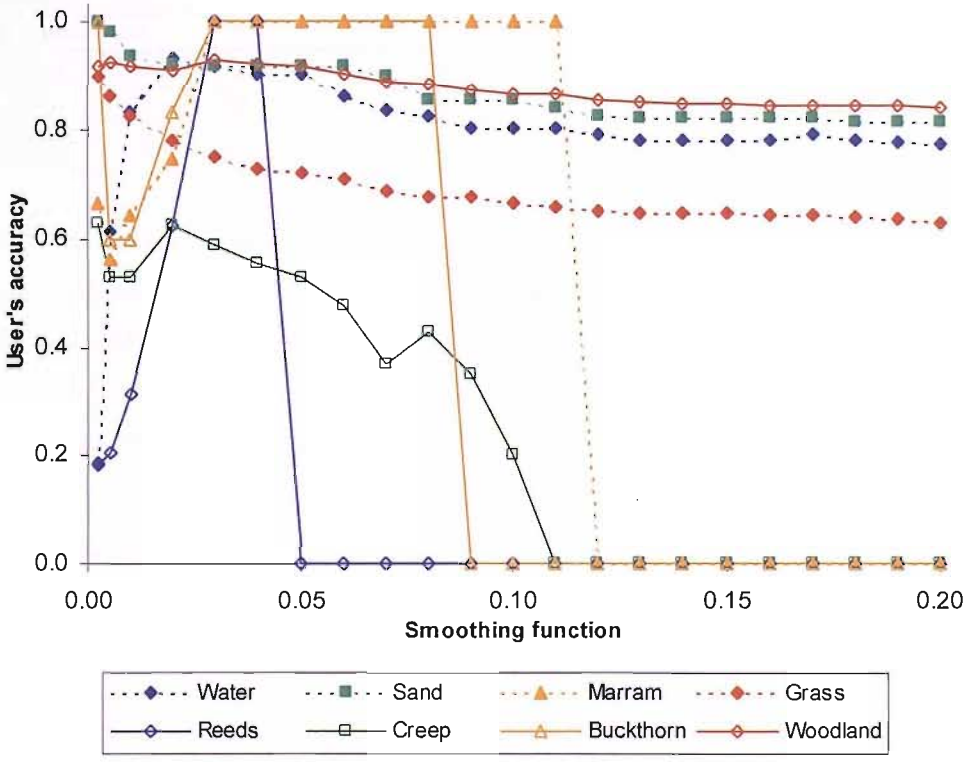


Figure 4.24 **PNN User's accuracy as a function of smoothing function.**

Table 4.3 Confusion matrix for PNN where h=0.02.

	Water	Sand	Marram	Grass	Reeds	Creep	Buckthorn	Woodland	Correct	Total	User's accuracy
Water	57	3	0	1	0	0	0	0	57	61	0.93
Sand	0	60	2	3	0	0	0	0	60	65	0.92
Marram	0	0	9	3	0	0	0	0	9	12	0.75
Grass	0	1	17	243	7	30	0	13	243	311	0.78
Reeds	0	0	0	3	5	0	0	0	5	8	0.63
Creep	0	0	1	25	4	52	0	1	52	83	0.63
Buckthorn	0	0	0	0	0	0	5	1	5	6	0.83
Woodland	0	0	1	18	0	0	2	219	219	240	0.91
Correct	57	60	9	243	5	52	5	219	650		
Total	57	64	30	296	16	82	7	234		786	
Producer's accuracy	1.00	0.94	0.30	0.82	0.31	0.63	0.71	0.94			0.827

Table 4.4 Confusion matrix for PNN where h=0.11.

	Water	Sand	Marram	Grass	Reeds	Creep	Buckthorn	Woodland	Correct	Total	User's accuracy
Water	57	10	2	2	0	0	0	0	57	71	0.80
Sand	0	53	3	6	0	1	0	0	53	63	0.84
Marram	0	0	1	0	0	0	0	0	1	1	1.00
Grass	0	1	23	265	13	77	3	19	265	401	0.66
Reeds	0	0	0	0	0	0	0	0	0	0	0.00
Creep	0	0	0	1	0	0	2	2	0	5	0.00
Buckthorn	0	0	0	0	0	0	0	0	0	0	0.00
Woodland	0	0	1	22	3	4	2	213	213	245	0.87
Correct	57	53	1	265	0	0	0	213	589		
Total	57	64	30	296	16	82	7	234		786	
Producer's accuracy	1.00	0.83	0.03	0.90	0.00	0.00	0.00	0.91			0.749

The PNN that had the smallest class specific thematic uncertainty RMSE was the one where the smoothing function equalled 0.02 (Figure 4.25). Above and below this value there was an increase in RMSE as the difference between the smoothing function value and 0.02 increased. This indicated that the PNN that was most suitable for predicting the thematic uncertainty for all classes was the PNN with a smoothing function of 0.02.

The difference that smoothing function made on the ability of PNNs to accurately classify certain classes and predict thematic uncertainty of those classes may be seen in Figure 4.26. In Figure 4.26a (h=0.02) the PNN predicted the class specific thematic uncertainty accurately. In Figure 4.26b (h=0.11) the PNN did not predict the class specific thematic uncertainty accurately. There were two outliers in this plot for Marram and Creep classes. Only one pixel was classified as Marram and this pixel was correct. Five pixels were classified as Creep, but all these pixels were incorrect. In addition, none of the accuracy assessment pixels were classified as Buckthorn and Reeds.

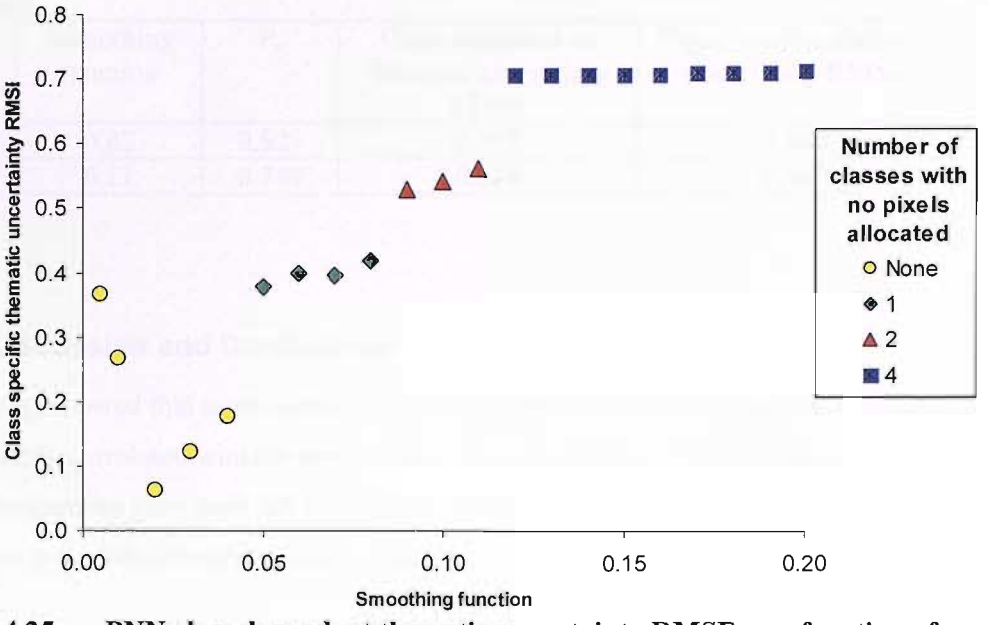


Figure 4.25 PNN class dependent thematic uncertainty RMSE as a function of smoothing function.

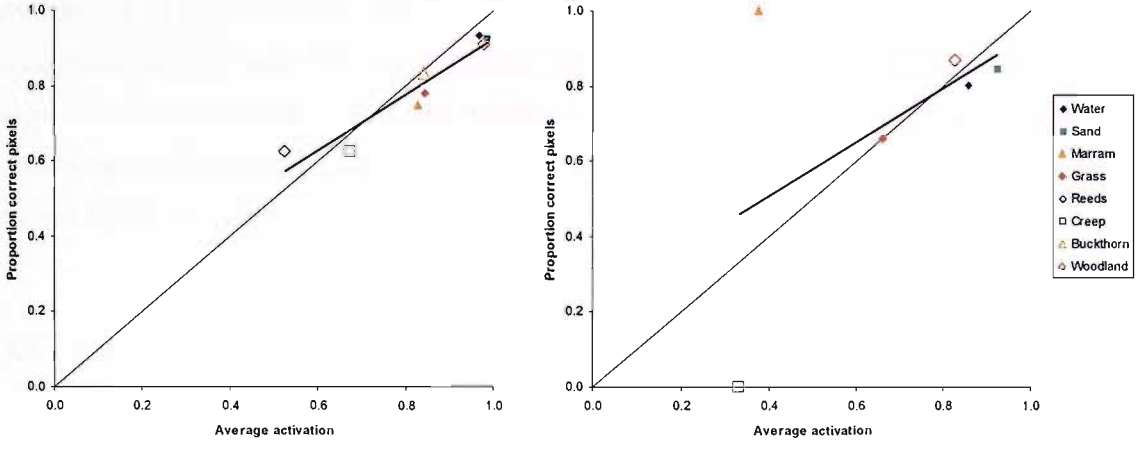


Figure 4.26 Proportion of correct pixels per class as a function of PNN activation. Thin line represents assumption of one to one relationship. Thick line represents linear regression. Buckthorn and Reeds were omitted from the graph when smoothing function equalled 0.11, as no pixels were allocated these classes.
a) Smoothing function = 0.02.
b) Smoothing function = 0.11.

Different PNNs were the most accurate at predicting class and thematic uncertainty (Table 4.5). The PNN where h=0.02 was the most accurate and predicted class specific thematic uncertainty most accurately. The PNN where h=0.11 predicted class independent thematic uncertainty most accurately.

Table 4.5 Most accurate PNN networks dependent on variable used to define accuracy.

Smoothing function	P_o	Class independent thematic uncertainty RMSE	Class specific thematic uncertainty RMSE
0.02	0.827	0.077	0.063
0.11	0.749	0.019	0.560

4.8 Discussion and Conclusions

The results showed that it was possible to derive thematic uncertainty measures that were significantly correlated with the proportion of correct pixels (at 95% confidence), using the output activations from both MLP and PNN neural networks. These thematic uncertainty measures provided estimates of the probability of correct allocation on a per-pixel basis.

However, the neural network that was most appropriate for one purpose was not necessarily the network that was most appropriate for another. For example, the PNN with a smoothing function of 0.02 had the largest overall accuracy (P_o = 0.827). This network also had a four times greater error when predicting class independent thematic uncertainty (RMSE=0.077) compared to the PNN with a smoothing function of 0.11 that had the smallest class independent thematic uncertainty RMSE (RMSE=0.019). The difference between the P_o of the PNN with the smallest RMSE (P_o = 0.749) and the PNN with the largest P_o was 0.078.

4.8.1 MLP

4.8.1.1 Overall accuracy

The results obtained in this study are broadly in agreement with previous studies on the effects of network variables on accuracy (Foody and Arora, 1997; Kavzoglu and Mather, 2003). Heuristics derived to maximise thematic accuracy have been explained in terms of ability of a network to model the complexity of the data, without becoming over trained or too complex and not able to generalise (Kanellopoulos and Wilkinson, 1997).

Accuracy increased as the number of nodes in the hidden layer increased, probably due to the network being able to model more complex distributions until it reached an optimum number of nodes (Figure 4.6) (Kavzoglu and Mather, 2003). Below this optimum, networks were less able to accurately model the complexity of the data. However, beyond the optimum point, the

increase in potential of a network to model complex distributions may have reduced the ability of a network to generalise and therefore accurately classify data other than that used in training.

Over the range of iterations used in this study for the MLP, overall accuracy generally increased with an increase in the number of iterations (Figure 4.7). Initially an increase in number of iterations resulted in an increase in accuracy. This was probably due to the network learning the complexities of the training data distributions. However, beyond an optimum number of iterations, MLPs learnt the training data too accurately, resulting in a decreased ability to generalise and therefore became less able to accurately classify data they had not seen before (Atkinson and Tatnall, 1997; Paola and Schowengerdt, 1997). It could be assumed from results that the optimum number of iterations to maximise the MLP's overall accuracy was not reached. However, by examining the maximum activation as a function of iteration it may be seen that though there was a general trend for accuracy to increase with number of iterations, the maximum accuracy of the networks tested was achieved with 2000 iterations (Figure 4.27). It was likely that the optimum number of iterations was dependent on the architecture of the MLP. A smaller network is less complex and therefore will take a smaller number of iterations to learn to model the distributions of the training data.

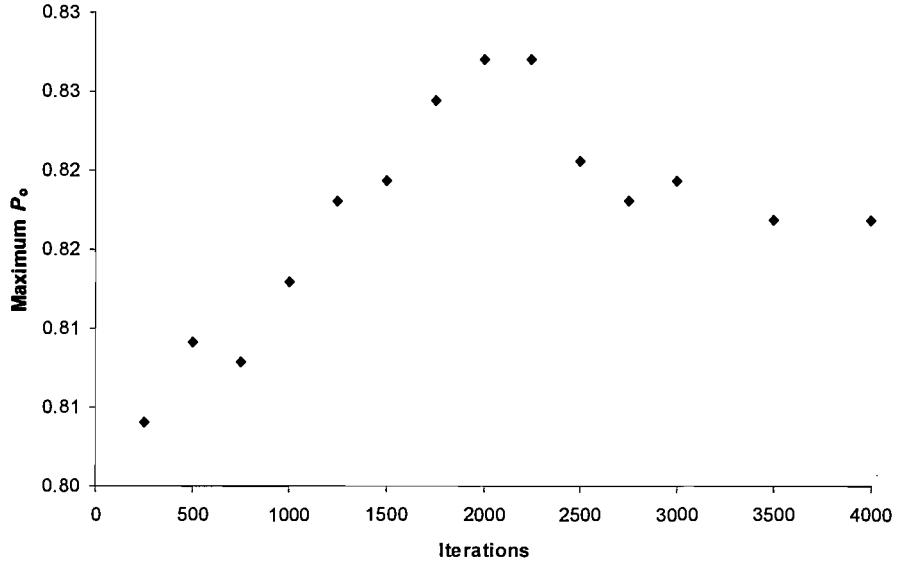


Figure 4.27 Maximum MLP P_o as a function of the number of iterations.

4.8.1.2 Thematic uncertainty

The results show that it was possible to derive measures that were significantly correlated (at 95% confidence), with per-pixel thematic uncertainty using the output activations from the MLPs. These measures provided indications of the probability of correct allocation. From the

results obtained it appeared that the ability of a network to accurately model thematic uncertainty was also sensitive to network variables.

Identifying the MLP network that modelled thematic uncertainty most accurately was difficult to define, as it depended on the definition of thematic uncertainty and there was no significant correlation between the class independent and class specific thematic uncertainty measures (at 95% confidence) (Appendix G; Table G.8). If thematic uncertainty was defined as being class independent, the most accurate network was different from the one that was most accurate if class specific thematic uncertainty was considered (Table 4.2). Neither of these networks matched the architecture for the most accurate network in terms of overall accuracy (Table 4.2). When there were five nodes in the MLP the network did not predict thematic uncertainty accurately. The MLP that predicted class independent thematic uncertainty most accurately had 10 nodes. As the number of nodes was increased beyond 10 there was a general decrease in the ability to model class independent thematic uncertainty (Figure 4.9).

The large class independent thematic uncertainty RMSEs obtained from the five node networks were likely to be due to the lack of complexity in the network. It may be argued that the network's ability to model the complexities of the data affected overall accuracy and prediction of thematic uncertainty. At network sizes larger than ten nodes, the ability of the network to model thematic uncertainty was reduced. This could be due to over training, where an increase in the ability of the network to model the complexities of the training data results in an increase in thematic uncertainty prediction error. However, in the case of the MLP, prediction of thematic uncertainty is a by-product of the classification process. The training process minimises the error of the overall classification and does not estimate posterior probabilities. The decision boundaries for an MLP are not based on a PDF as with a Bayesian approach to classification. To position the decision boundary between classes in a hard classification so that overall accuracy is maximised there is no need for a probabilistic output. Classification accuracy does not necessarily correlate with the ability of a network to predict thematic uncertainty. A network may produce a large accuracy value, but the activations may tend towards the extremes of 0 and 1. When the relationship between P_o and normalised entropy was examined (Figure 4.16) there was a negative correlation, and so as accuracy increased, entropy decreased, indicating that activations tended towards extremes. This is an indication that as the networks were trained they became more confident in their outputs, though this increased confidence did not result in a greater ability to predict uncertainty.

If activations tend towards 0 or 1 for a classification containing thematic errors, thematic uncertainty is less likely to be modelled accurately, as the full range of probabilities will not be represented. There would be an exception if pixels were correctly allocated, as probabilities could be represented as certainties (0 or 1). To model thematic uncertainty accurately, the network has to be able to output activations over the full range of probabilities.

Entropy is a measure of uncertainty, and will be small when activations tend to 0 and 1. If the entropy was small then it may be assumed that the network was not able to represent thematic uncertainty unless pixels were correctly allocated, as it did not tend to output activations away from 0 and 1. If this were the case then a network with small entropy would be less able to represent uncertainty than one with large entropy. There was a significant but weak relationship between class independent thematic uncertainty RMSE and entropy (R^2=0.145, F=19.8, p<0.001) (Figure 4.28) indicating that the ability of a network to model thematic uncertainty was only partially a function of the ability of the network to output activations that were not close to 0 or 1. Another interpretation would be that large entropy represented large uncertainty in the thematic output and did not necessarily allow the network to model uncertainty accurately. This was borne out as, relative to the other MLPs, the networks with 5 nodes had large entropy values, but large class independent thematic uncertainty RMSE and small accuracy values. This complex relationship meant that a relationship between P_o and class independent thematic uncertainty RMSE was unlikely and this was borne out by the results (R^2=0.012, F=2.4, p=0.126).

The thematic uncertainty results were made more complex when the class specific thematic uncertainty as a function of network variables results was examined. There was no significant correlation between class specific thematic uncertainty RMSE and number of nodes (at 95% confidence) (Figure 4.12). However, there was a negative log-linear correlation between class specific thematic uncertainty RMSE and the number of iterations (Figure 4.14) (R^2=0.922, F=155.5, p<0.001). The optimum network for predicting class specific thematic uncertainty (30 nodes, 3500 iterations) was not the optimum network for prediction of either class independent thematic uncertainty or overall accuracy (Table 4.2).

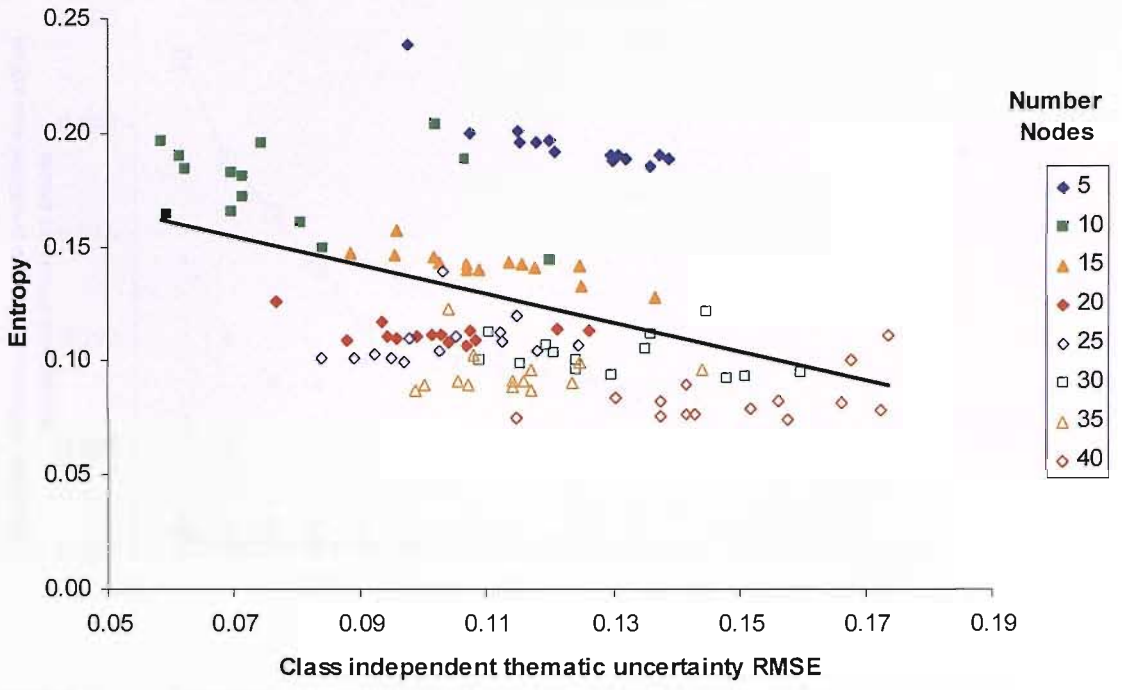


Figure 4.28 MLP entropy as a function of class independent thematic uncertainty RMSE.

The accuracy of thematic uncertainty measures derived from the MLP was partially a function of network variables such as number of iterations and number of nodes. However, these heuristics should not be applied to other datasets, as variation in accuracy of thematic uncertainty measures is also dependent on variables other than network architecture. These include the land cover type being classified, the number of classes, the datasets used.

The possibility for the discrepancy between class independent and class specific thematic uncertainty RMSE trends was likely to be due to the components of the RMSE value calculated. Creep was the class that made up the greatest proportion of class specific thematic uncertainty error for most iterations (Figure 4.29). The error in predicting the thematic uncertainty associated with this class decreased as the number of iterations increased. From the ground data it may be seen that the Creep class made up less than 11% of the total number of pixels (82 pixels of 786; Table 4.3). In terms of class specific thematic uncertainty this class had a disproportionate effect compared to its relative area.

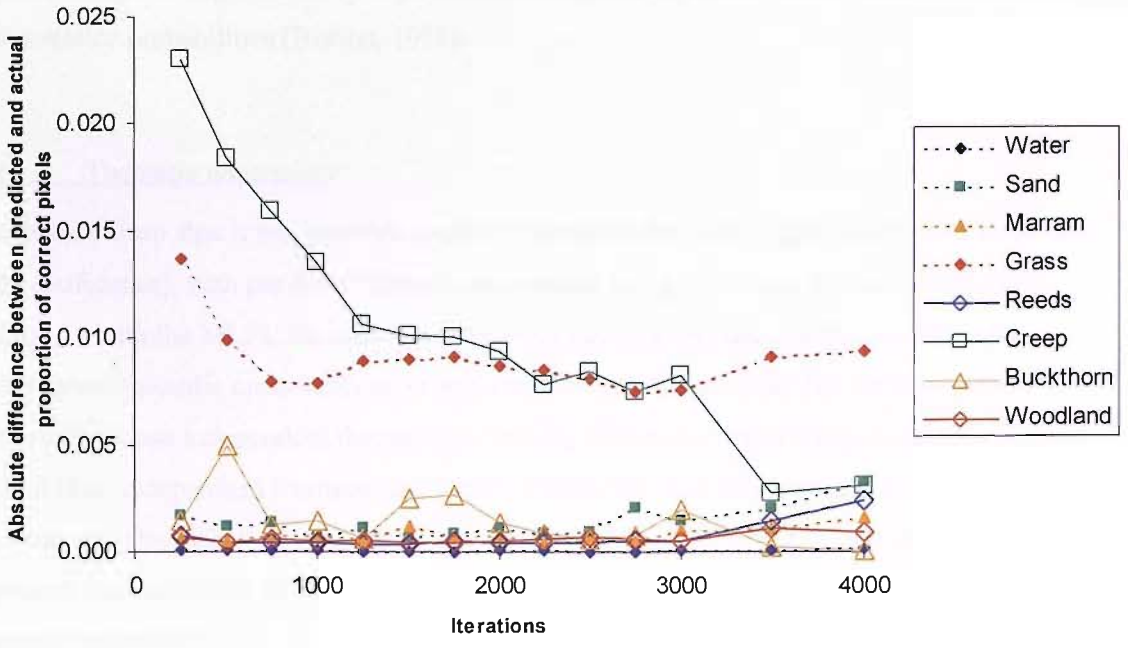


Figure 4.29 Absolute difference between predicted and actual proportion of correct pixels on a per-class basis.

The variation in the range of overall accuracy and errors in predicting thematic uncertainty varied a great deal in the networks tested (Table 4.2). The minimum overall error derived from P_o was 0.173, 37% less than the maximum of 0.274 (Appendix G). There was a 66% variation in class independent thematic uncertainty RMSE and a 54% variation in class specific thematic uncertainty RMSE (Appendix G). The error when predicting thematic uncertainty varied more than when predicting class. This may have been due to the error measures during training being derived from thematic accuracy rather than thematic uncertainty. The training process therefore minimised thematic error rather than the error associated with predicting thematic uncertainty.

4.8.2 PNN

4.8.2.1 Overall accuracy

The PNN P_o was dependent on smoothing function. Below the optimum value for accuracy (h=0.02), P_o decreased rapidly with small changes in smoothing function. This was likely to be due to the PDF being a noisy representation of the training data (Bishop, 1995). The network was able to classify the training data very accurately, but could not generalise to the rest of the dataset. For values of smoothing function above 0.02, P_o decreased slowly with an increase in smoothing function. This was likely to be due to the smoothing function and so the training

kernels being too large, resulting in an over-smoothed PDF, which would be unrepresentative of the posterior probabilities (Bishop, 1995).

4.8.2.2 Thematic uncertainty

The results show that it was possible to derive measures that were significantly correlated (at 95% confidence), with per-pixel thematic uncertainty using the output activations from the PNNs. As with the MLPs, the most accurate PNN network and the one that predicted class independent thematic uncertainty most accurately were not the same. The PNN with h=0.11 had the smallest class independent thematic uncertainty RMSE and there was a large difference in P_o and class independent thematic uncertainty RMSE between this network and the most accurate one (h=0.02, P_o = 0.827, RMSE = 0.077; h=0.11, P_o = 0.749, RMSE = 0.019). This indicates that the ability of a PNN to classify accurately was not reflected in its ability to predict thematic uncertainty.

However, the class independent thematic uncertainty RMSE values did not fully describe thematic uncertainty prediction using the PNN, as they did not provide per-class information and, therefore, may not have fully represented the thematic uncertainty prediction error. By examining the confusion matrices it was found that in all but six of the PNNs tested, at least one class was not represented in the accuracy assessment data, as no pixels were allocated to the class. Only in the classifications with a smoothing function between 0.0025 and 0.04 were all the classes represented.

For the PNN with the smallest class independent thematic uncertainty RMSE (h=0.11), none of the pixels in the accuracy assessment were classified as Reeds or Buckthorn (Table 4.4), and the Marram class was only represented by one classified pixel, even though there were 30 ground data pixels in the accuracy assessment data. There were only five Creep class pixels, all of which were misclassified, out of 82 in the accuracy assessment ground data. The PNN with h=0.11 would not represent land cover accurately as these classes would be either under-represented or not represented at all in the final classification. This was likely to be due to an over-smoothed class conditional distribution resulting from too large a smoothing function. The estimated prior probabilities for the Reeds, Buckthorn, Marram and Creep classes were all small relative to the other vegetation classes, Woodland and Grass (Table 4.1). These small prior probabilities linked with an over-smoothed class conditional distribution would result in misclassification as a spectrally similar class, but with a larger prior probability. Most of the

misclassifications resulted from the Reeds, Buckthorn, Marram and Creep classes being misclassified as Woodland and Grass (Table 4.4).

By examining the thematic uncertainty on a per-pixel per-class basis, it may be seen that the RMSE values based on all activations from the network did not necessarily represent uncertainty accurately. Only six of the PNNs resulted in at least one pixel for each class during the accuracy assessment stage. Of these, the PNN that represented class specific thematic uncertainty most accurately on a per-pixel per-class basis had h=0.02, which coincided with the largest overall accuracy. From the plot of per-class average activation against the proportion of pixels correctly classified for the PNN with h=0.02, it may be seen that the network modelled per-pixel uncertainty accurately for all classes (Figure 4.26). A linear regression of actual proportion of pixels correctly classified as a function of the predicted value assuming that this was the average activation showed that there was a significant correlation (R^2= 0.886, F=55.5, p<0.001).

4.8.3 Comparing the MLP and PNN

Both PNN and MLP networks were able to model thematic uncertainty on a per-pixel basis. For both types of network, there did not appear to be a link between the thematic accuracy of a network and the thematic uncertainty outputs. Though selecting the most accurate network or the one that modelled thematic uncertainty most accurately was relatively simple, determining network suitability for change detection was more difficult. This was because the requirements of error prediction are different for a simple mapping exercise using the output from a classification and monitoring such as change detection where two or more classifications are combined. If a land cover map is required, the most accurate network was likely to be most suitable. However, for change detection there was also a requirement for accurate prediction of uncertainty, to estimate the probability of error of the output. In simple terms the optimum network should have the largest accuracy and smallest class independent and class specific thematic uncertainty RMSEs. These conditions did not coincide for the MLP or the PNN networks. Therefore, a network used for change detection where output errors are modelled should be selected carefully to reduce both thematic error and thematic uncertainty error. This is likely to result in a compromise between small thematic uncertainty errors and small thematic errors.

It was shown that care had to be taken when examining the class independent and class specific thematic uncertainty RMSE values as they did not necessarily fully represent thematic

uncertainty accuracy. Using the class independent thematic uncertainty RMSE, there was no indication of how accurately class specific thematic uncertainty was represented.

It was likely that the PNN most suited for change detection in which change from all eight classes was a possibility was the one with h=0.02. This was because of all the PNNs tested this network had the largest overall accuracy, was able to represent all classes within the classification and could predict the probability that a given class was correct with the least error.

During the training process MLPs move the boundaries between classes in order to minimise error (Bishop, 1995). This indicates that MLP activations only represent thematic uncertainty as a by-product of the classification process and cannot be assumed to represent posterior probabilities. PNNs use the posterior probability to determine the most probable class and the decision boundaries between classes. As PNNs generate the posterior probability, the outputs may be directly interpreted as thematic uncertainty. This indicates that PNNs should be more suitable than the MLPs for deriving thematic uncertainty on a per-pixel basis. However, the results obtained in this study do not necessarily indicate this, as the PNN that represented thematic uncertainty most accurately (Table 4.5), had greater thematic uncertainty errors than the MLP that represented thematic uncertainty most accurately (Table 4.2). Though the most accurate PNN (h = 0.02) (Table 4.5) represented thematic uncertainty less accurately than some of the MLPs (for example 10 nodes, 1000 iterations), when overall accuracy, class independent thematic uncertainty and class specific thematic uncertainty were taken into account no single PNN or MLP was most accurate (Table 4.2).

To simplify the study a single network type was selected for use in later chapters. The PNN's outputs are posterior probabilities, unlike the MLP for which the outputs need to be interpreted as posterior probabilities. The PNN also produces a stable output, as the output will be the same for the same training inputs and network variables. The accuracy of MLPs is dependent on starting conditions, as well as network variables. Small changes in the initial conditions prior to training can result in large differences in the final ability of the network to classify even with the same input training data (Kanellopoulos and Wilkinson, 1997; Paola and Schowengerdt, 1997). For these reasons the PNN was used in later chapters for change detection.

Care has to be taken when using indicators of the accuracy of thematic uncertainty measures such as the class independent thematic uncertainty RMSE. These indicators may not fully represent the ability of a network to predict thematic uncertainty. Within this study it was seen that global measures of uncertainty such as the class independent thematic uncertainty RMSE

can give an incorrect impression of how thematic uncertainty was represented by output activations. The confusion matrix should be studied to determine whether all classes were accurately classified, particularly in the case where some of the classes had a very small prior probability and so were not accurately represented in the accuracy assessment stage.

4.8.4 Summary

In this chapter thematic uncertainty measures were produced that could be used to map the spatial distribution of thematic errors in classifications for a model of change detection. These thematic uncertainty measures were derived from the output of MLP and PNN classifiers and provided indications of the thematic uncertainty for every class at every pixel. Other classification methods may be used to provide per-pixel thematic uncertainty (for example de Bruin and Gorte, 2000; McIver and Friedl, 2001) and, therefore, used as input to per-pixel change detection models. The outputs of PNNs are the posterior probability of correct allocation for every class at every pixel and therefore PNNs are more suitable for a probabilistic study than MLPs that do not output probabilities.

Though there were significant correlations between overall accuracy and thematic uncertainty as a function of network variables for MLPs, heuristics for deriving the most suitable network for change detection should not be made on the basis of these trends, as the variables were not highly correlated and are likely to be very data dependent. This means that the most effective method of determining a network to use should be based on empirical testing rather than using pre-determined heuristics.

5 Combining thematic and misregistration uncertainty using synthetic data

5.1 Introduction

The purpose of this study was derive operational methods of determining land cover change using remote sensing that would minimise the error associated with change detection. To do this a model of land cover change that predicted the uncertainty associated with the process of determining thematic change on a pixel by pixel basis was derived. Assuming that a hard thematic model can represent the real world, the errors associated with change detection are due to thematic errors in the classifications and misregistration between images that were classified. Determining per-pixel change therefore required the per-pixel prediction of thematic error and misregistration errors between classification data. As it is effectively impossible to determine error on a per-pixel basis for a large area, measures of the probability of error or uncertainty were required. The use of geometric and thematic uncertainty measures was tested to determine whether these measures could increase the accuracy of change detection.

5.2 Use of thematic uncertainty measures in change detection

In the previous chapters, per-pixel measures of thematic and misregistration uncertainty were derived. If these measures were to be of use in per-pixel change detection it must be possible to combine them to generate a per-pixel prediction of change. If per-pixel measures of the probability of change were to be generated then the misregistration and thematic error measures must be probabilistic.

In Chapter 3 an overall misregistration error function, μ, was derived (Section 3.6.6). However, μ was not suitable as a measure of misregistration uncertainty, as it was a single averaged value for each pixel and was not a probabilistic output. To provide a per-pixel change uncertainty output from the thematic and geometric uncertainty measures, it was necessary to preserve the probabilistic nature of the input thematic and misregistration models. The geometric error model that was used in the generation of the averaged misregistration model derived was probabilistic. The output of the geometric error model consisted of a series of geometric error vectors for every pixel with a probability associated with each vector (Figure 3.22). This uncertainty model was suitable for use in per-pixel change detection, as it provided a local model of geometric uncertainty. When the geometric uncertainty models are combined for two classifications at two times, they can provide probabilistic measures of misregistration.

In Chapter 4 outputs from MLP and PNN classifiers were used to derive per-pixel measures of thematic uncertainty (Figure 5.1). These measures could be directly used in a change uncertainty model, as they are probabilistic or could be interpreted as probabilistic and provide a thematic uncertainty measure for every class at every pixel (Figure 5.1).

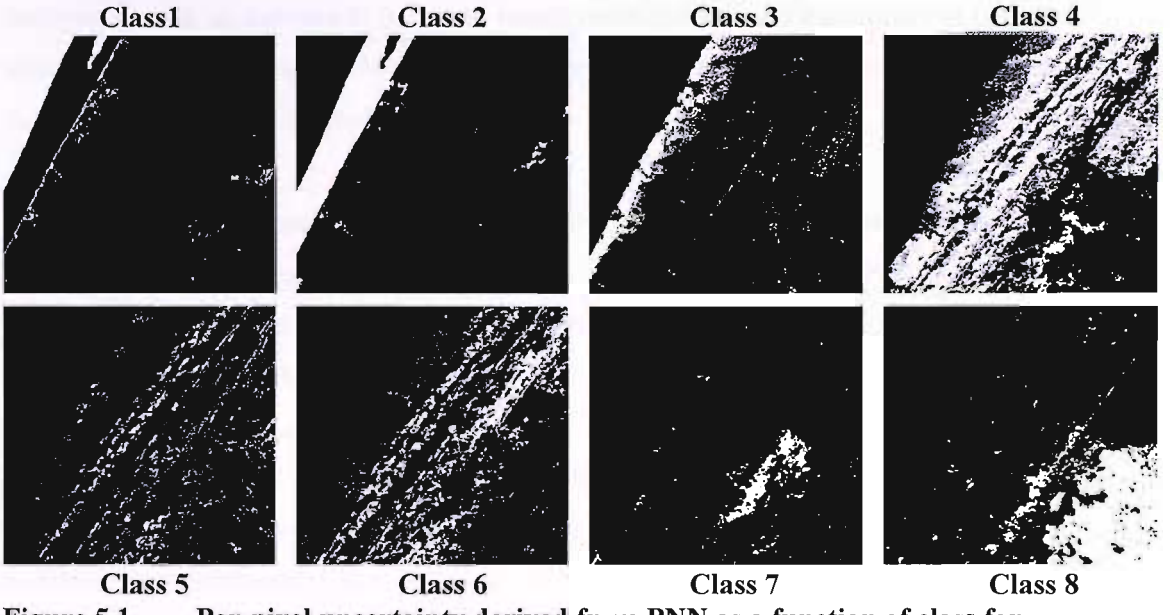


Figure 5.1 Per-pixel uncertainty derived from PNN as a function of class for subsection of 2002 Ainsdale data. White is large probability, black is small probability.

5.3 Fuzziness and thematic uncertainty

Errors in change detection are particularly likely to occur at boundaries between classes (Dai and Khorram, 1998). This is partially if the boundary position is incorrect due to misregistration. However, where there are fuzzy boundaries or ecotones it is difficult to position the boundary accurately (Brown, 1998; Fortin *et al.*, 2000; Molenaar and Cheng, 2000). Where pixels are mixed the hard thematic model is limited (Fisher, 1997; Foody and Boyd, 1999). One option would be to use fuzzy classification (Bastin, 1997; Foody, 1997; Foody and Boyd, 1999; Maselli, 2001) in change detection. However, the aim of this study was to use hard classification for thematic change detection, as potential users of the output are likely to require a traditional thematic map (Section 4.2).

If the thematic uncertainty measures provided an indication of the uncertainty due to mixed pixels, then this option would account for errors in the classification as well as the fuzziness (Foody, 2000). However, thematic uncertainty and fuzziness are inherently different concepts (Ricotta, 2004). Fuzziness is partial membership to a series of classes, whereas thematic

uncertainty is a description of the error associated with the probability of membership to a series of classes. In the context of hard classification, fuzziness may be considered a subset of uncertainty. As the amount of mixing within a pixel increases, the probability that a given hard class is correct is reduced due to confusion of the classes in feature space. It was therefore likely that the thematic uncertainty measures derived in Chapter 4 would partially account for the multiple class membership of a pixel as well as the uncertainty. In the case of pixels that contain multiple classes, an increase in fuzziness would result in increased uncertainty in the correct allocation. It follows that the thematic uncertainty measures derived could account for fuzziness in terms of uncertainty in a hard classification.

In this study it was not possible to test whether the measures derived were both fuzzy and probabilistic, as fuzzy accuracy assessment data were not acquired. The PNN outputs posterior probabilities as with the ML classifier, though the PNN is non-parametric unlike the ML. Previous studies have shown that the ML output can be used to predict multiple class membership (Foody *et al.*, 1992; Bastin, 1997). Therefore, posterior probability can be used to model fuzziness and so the assumption was made that there was a link between fuzziness and PNN output and that the thematic uncertainty measures derived were likely to account for the multiple class membership, as well as uncertainty.

5.4 Combining geometric and thematic uncertainty measures

To combine the geometric and thematic uncertainty models derived in a per-pixel change uncertainty model the following approach was taken (Figure 5.2):

1. For every geometric error vector of a given pixel, the thematic uncertainty values were applied to the position of the pixel represented by the geometric error vector. As the geometric error vector was not necessarily an integer value, there were likely to be four pixels that a thematic uncertainty value was applied to due to a single geometric error vector (Figure 3.22; Figure 5.2). The thematic uncertainty value was modified by the proportion of a pixel the geometric error vector applies to and by the probability of the vector.
2. Stage 1 was repeated for every class.
3. The new thematic uncertainty was normalised, so that the uncertainty for every pixel equalled one.

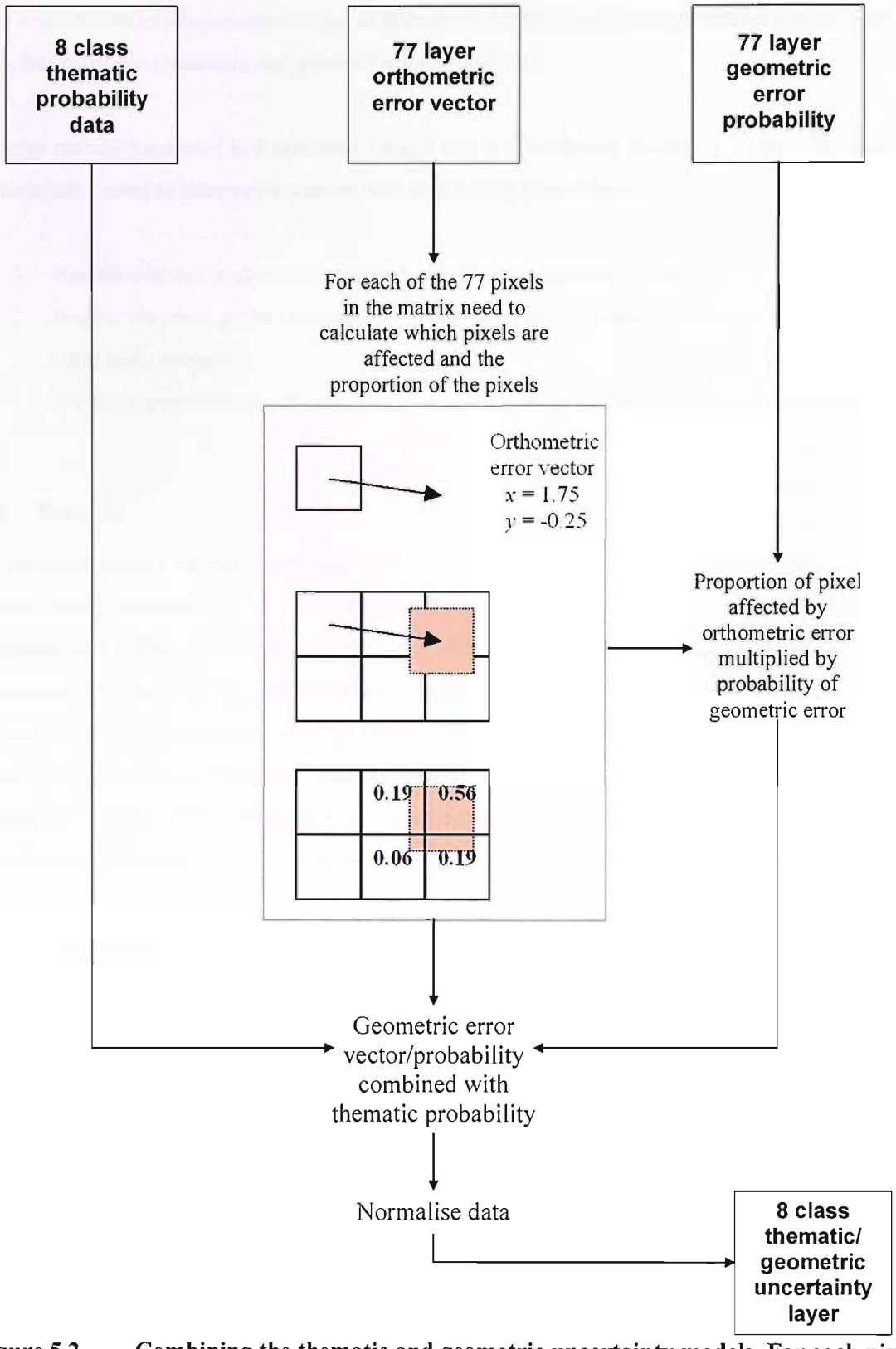


Figure 5.2 Combining the thematic and geometric uncertainty models. For each pixel there are up to 77 orthometric error vectors and associated geometric error probabilities. This represents the maximum size of the geometric error matrix in Figure 3.21.

The new uncertainty layer was similar in appearance to the thematic uncertainty layer (Figure 5.1), but combined thematic and geometric uncertainty.

Change may be predicted in a number of ways, but in this chapter the ability of the combined uncertainty model to determine change was assessed in three ways:

1. Predict whether a given pixel had changed or not between t_1 and t_2.
2. Predict the most probable class of a given pixel at two times, and determine whether the class had changed.
3. Predict the probability or magnitude of change in land cover class between t_1 and t_2.

5.5 Method

To compare change detection methods, 256 pixel by 256 pixel synthetic images were generated with a spatial resolution of 1 m, as with the Ainsdale CASI data. Each image was treated as a complete and 100% accurate ground data set. For each change scenario one image was generated for t_1 and one for t_2. The images contained two classes. Class 1 was a square of variable size in the centre of the image, Class 2 the area surrounding the square (Figure 5.3). For t_1, the size of the classes was kept constant, but the size of the boundary between them was changed. The sizes of the boundary width used were 0 m (a hard boundary), 5 m, 10 m, 15 m, and 20 m. The mixing in the boundaries was assumed to be linear with distance.

The t_2 images were generated with the same variations in boundary fuzziness as for t_1. For each image at t_1, a no change image was generated for t_2. Images were also generated with varying amounts of change between t_1 and t_2 (Figure 5.3). The boundary between the two classes was moved varying distances in the t_2 image to generate differences of 5 m, 10 m, 20 m, 30 m and 40 m between t_1 and t_2. In every case the change was a reduction in the size of Class 1. These resulted in thirty change scenarios with varying amounts of change and varying widths of fuzzy boundary.

For each image representing actual ground conditions, thematic and geometric errors were applied to derive a synthetic image representing a classification derived from remotely sensed data.

Geometric error was applied assuming angular acceleration between $1°s^{-1}s^{-1}$ and

$10°s^{-1}s^{-1}$ for t_1 and between $1°s^{-1}s^{-1}$ and $5°s^{-1}s^{-1}$ for t_2. The average error was derived using Equation 2.11 and a random error between –1 and +1 pixel was applied to the average. The values used to generate the geometric error matched values observed in actual data. At the first pixel in the flight line the direction of the error vector was along the flight path. At pixel 64 in the flightline, the error vector was increased 90°. This was repeated at pixels 128 and 194. The variation in the direction of the error vector was applied so that misregistration direction varied across the imagery.

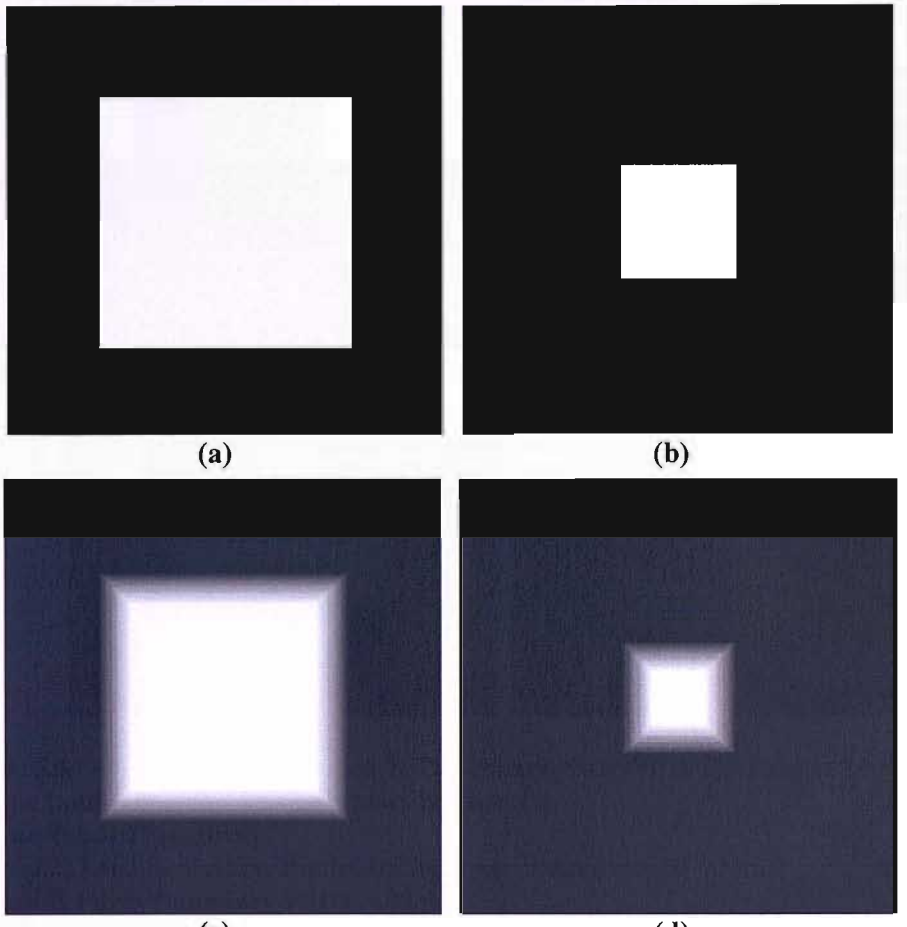
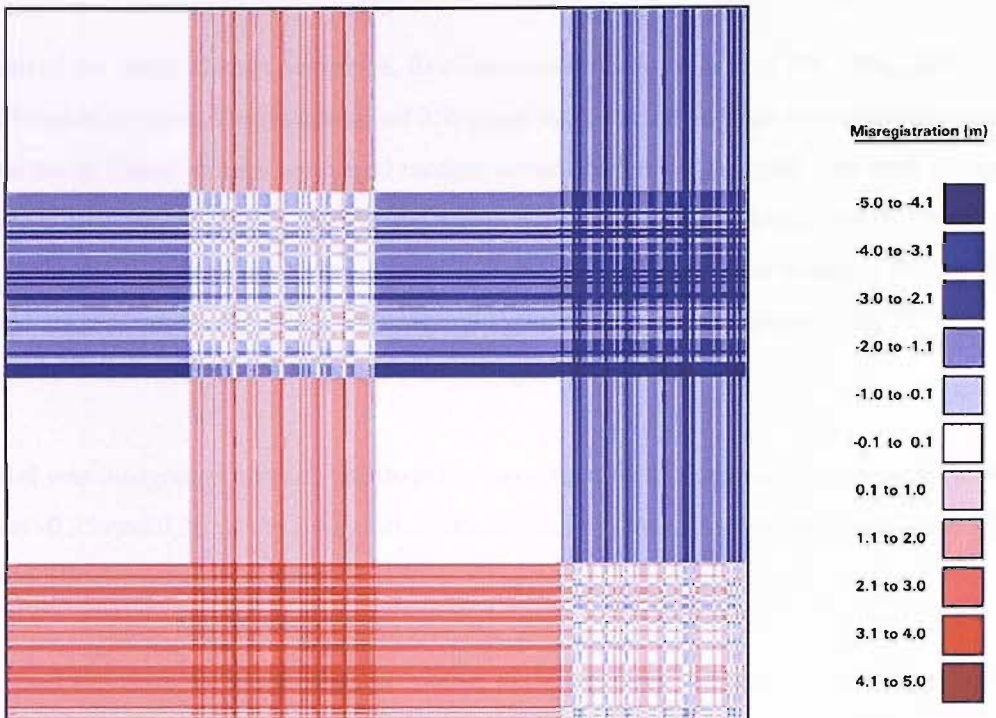
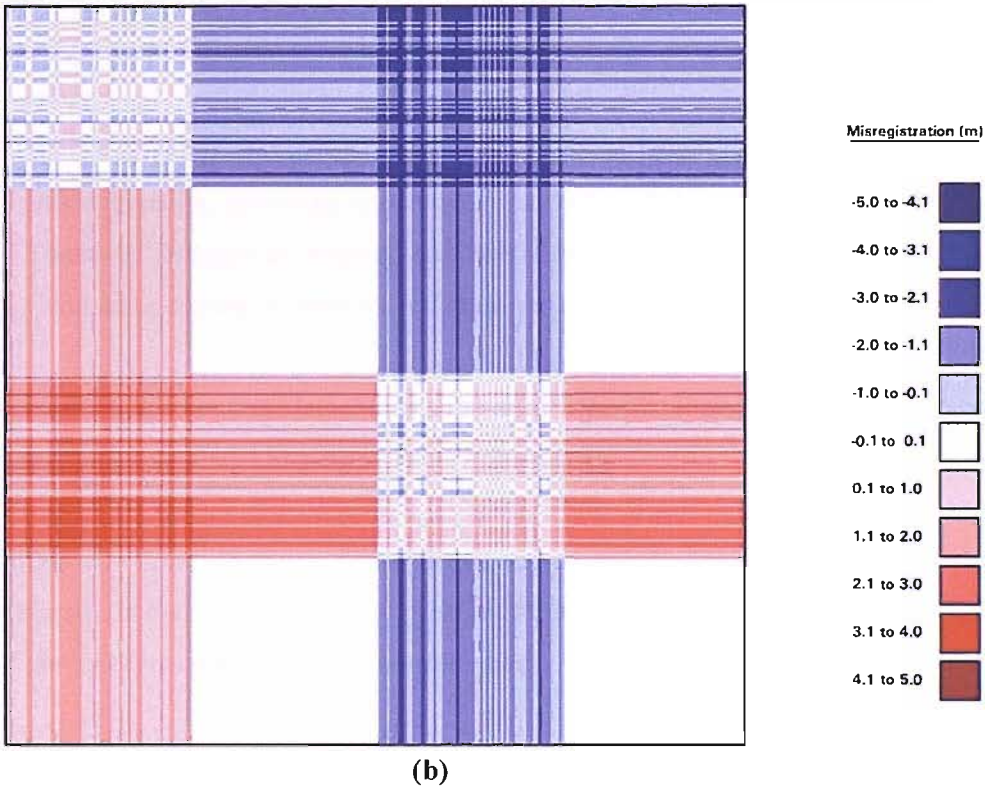


Figure 5.3 Synthetic data class 1. The change distance is assumed to be the distance that the boundary between the classes has moved.
a) Time 1, hard boundary.
b) Time 2, hard boundary. Boundary between classes moved 40 m.
c) Time 1, Fuzzy boundary width = 20 m.
d) Time 2, Fuzzy boundary width = 20 m. Boundary between classes moved 40 m.

For the t_1 images the assumed flight direction was south, for t_2 flight direction was west. The variation in flight direction was applied so that misregistration would vary across the image in two dimensions. Topographic effects were not taken into account in order to simplify the synthetic data model. The geometric error applied resulted in the misregistration in Figure 5.4



(a)



(b)

Figure 5.4 Misregistration between synthetic images. As the flight directions of the images at t_1 and t_2 were different by 90^O and the geometric error vector was varied across the image misregistration varies in two dimensions across the image.
a) Misregistration in x direction.
b) Misregistration in y direction.

For each of the thirty change scenarios, five thematic error scenarios of 5%, 10%, 20%, 30% and 40% were applied. Two randomised 256 pixel by 256 pixel images were derived, one for t_1 and one for t_2. These images contained random values between 0 and 100. For each of the thematic error scenarios (between 5% and 40%), a pixel was assumed to be correct if the pixel value of the randomised image was greater than the thematic error. For example for the 20% thematic error scenario, all pixels with a value in the random image greater than 20 were assumed to be correct. All other pixels were assumed to be incorrect.

If a pixel was designated correct, the thematic uncertainty was assumed to be a random value between -0.25 and 0.25 of the fuzzy value, approximately four times the most accurate PNN class specific thematic uncertainty RMSE in Chapter 4 (Table 4.5).

If the pixel was designated incorrect, a random thematic value between 0.5 and 1 was applied to the incorrect class. For example if the class of a pixel that had been designated incorrect was Class 1 in the original data then a value between 0.5 and 1 was applied to Class 2. This resulted in a series of images containing thematic and geometric errors (Figure 5.5).
The following change detection models were compared:

1. No Uncertainty: hard classifications and no geometric or thematic uncertainty.
2. Misregistration: geometric uncertainty was derived, but not thematic uncertainty.
3. Thematic: thematic uncertainty was derived but not geometric uncertainty.
4. Combined: combined thematic and geometric uncertainty.

The geometric uncertainty model and thematic uncertainty models were merged using code written in Interactive Data Language (IDL) (Appendix I).

The confusion matrices of the change models were compared using two models with either two or four classes. In the two-class change problem the classes were Change and No change. The class of the pixel was assumed to be Change when:

$$(\text{Class } 1_{t1} - \text{Class } 1_{t2}) >= 0.5$$

The class of the pixel was assumed to be No change when:

$$(\text{Class } 1_{t1} - \text{Class } 1_{t2}) < 0.5$$

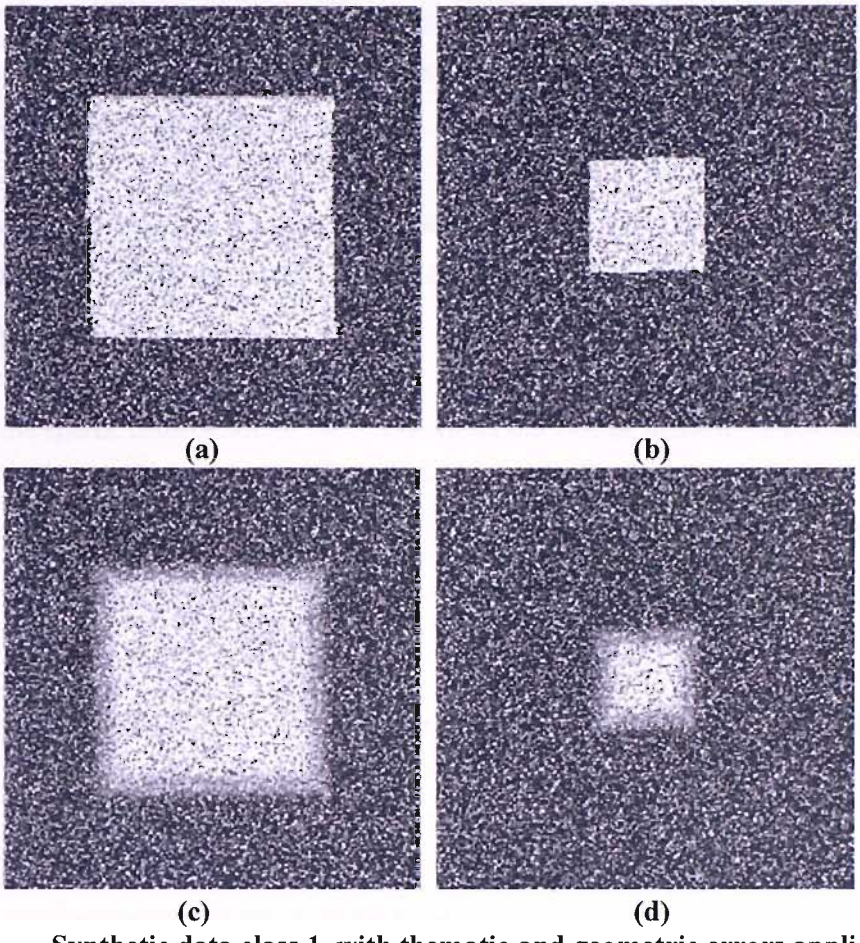


Figure 5.5 Synthetic data class 1, with thematic and geometric errors applied
a) Time 1, hard boundary.
b) Time 2, hard boundary. Boundary between classes moved 40 m.
c) Time 1, Fuzzy boundary width = 20 m.
d) Time 2, Fuzzy boundary width = 20 m. Boundary between classes moved 40 m.

In the four-class change problem there were the following classes:

1. No change, Class 1
2. Change, Class 1 to Class 2
3. Change, Class 2 to Class 1
4. No change, Class 2

To calculate which of the four classes each pixel belonged to, the class for t_1 and t_2 was derived from the class with the largest thematic uncertainty. As this was effectively a method of hardening a classification, the Thematic and No uncertainty models gave the same result. Using the four class set would enable judgements to be made as to how accurately the models could detect change and predict what change was taking place

The RMSE between the actual and predicted thematic change vector magnitudes was also compared for the four models. This enabled a comparison to be made of how accurately each model predicted the thematic change vectors.

The error statistics were generated from the whole of an image and therefore a whole population. This means that significance values did not need to be generated and any differences were effectively significant.

To examine the spatial distribution of the errors within the change detection methods the change images were observed when the thematic error was 20%. This value was used as it was the closest value to the thematic error of the most accurate PNN in Chapter 4 (Table 4.5).

5.6 Results

All the results from the synthetic data are in Appendix J.

5.6.1 Thematic error

As thematic error of the input layers increased, the overall accuracy of the change detection model decreased (Figure 5.6; Figure 5.7). The exceptions to this trend were the Misregistration and Combined models when no change had taken place using the two-class Change/No change model (Figure 5.6). The general trend of decreasing change detection accuracy with an increased thematic error was expected, as an increase in the input errors was likely to lead to an increase in the output errors. However, the near constant overall accuracy with an increased thematic error for the Misregistration and Combined uncertainty models was unexpected. When the thematic error was 20% both uncertainty models reduced the predicted change close to 0% (Table 5.1).

In the two-class problem, change was only assumed when the difference vector between the two times was greater than 0.5. As thematic errors were introduced randomly rather than in a clumped form it was likely that the majority of pixels containing thematic error would be surrounded by pixels that did not. In a pixel that contained thematic error, the geometric uncertainty model was likely to result in thematic error being smoothed below the 0.5 threshold by those correct pixels around it. Using the geometric uncertainty model each pixel was likely to be strongly affected by the pixels surrounding it, as the geometric error matrix used values from up to 76 of the surrounding pixels (Figure 3.21).

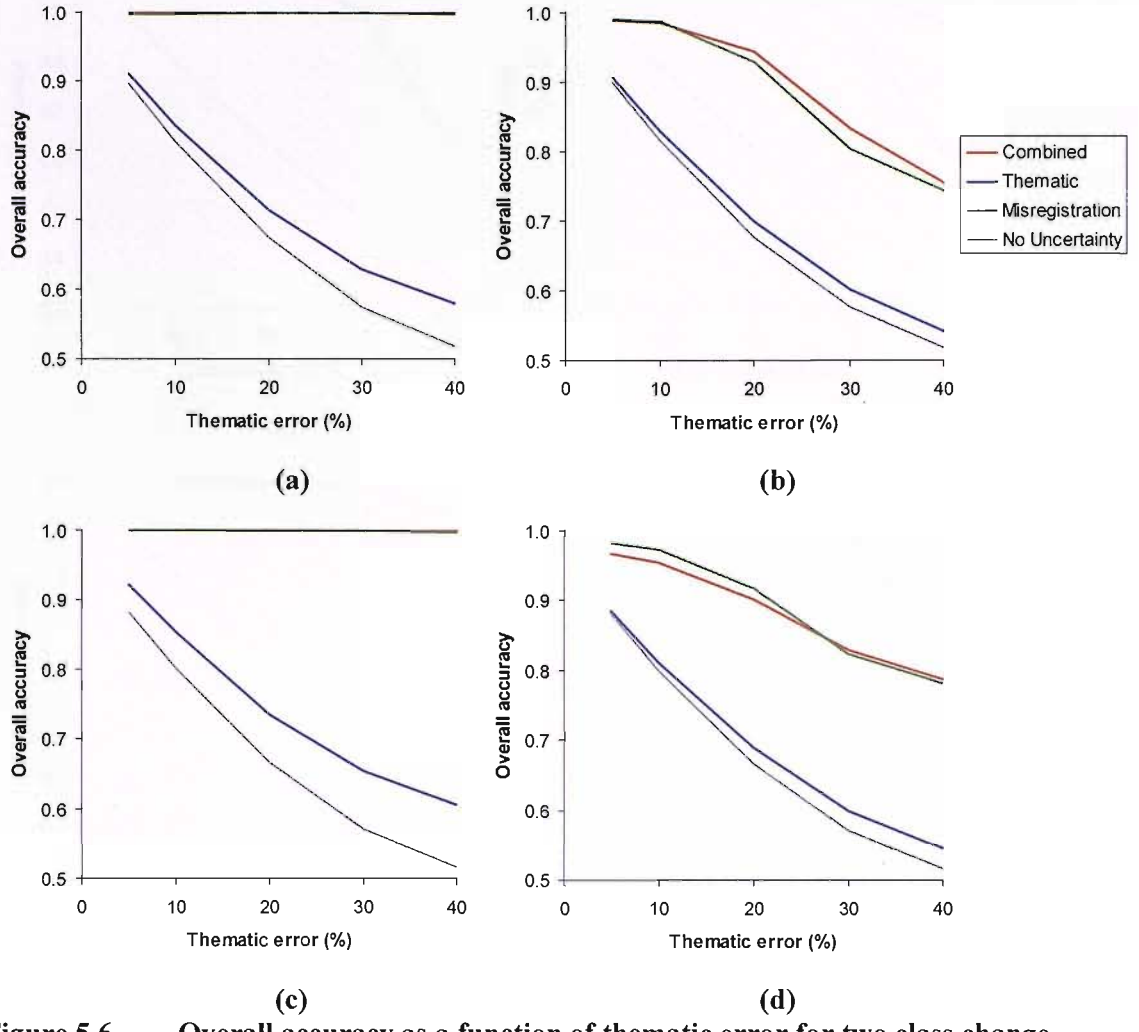


Figure 5.6 Overall accuracy as a function of thematic error for two class change model.
a) Time 1, hard boundary.
b) Time 2, hard boundary. Boundary between classes moved 40 m.
c) Time 1, fuzzy boundary width = 20 m.
d) Time 2, fuzzy boundary width = 20 m. Boundary between classes moved 40 m.

In the four-class problem, the No Uncertainty change detection method had a close to linear negative relationship with thematic error for all change scenarios (Figure 5.7). At a thematic error of less than 20%, the Combined and Misregistration change methods had a close to linear relationship with overall accuracy and overall accuracy values were greater than 97% (Figure 5.7). With a thematic error of greater than 20%, overall accuracy decreased sharply with an increase in thematic error, with the Misregistration method having the greatest decrease in overall accuracy (Figure 5.7).

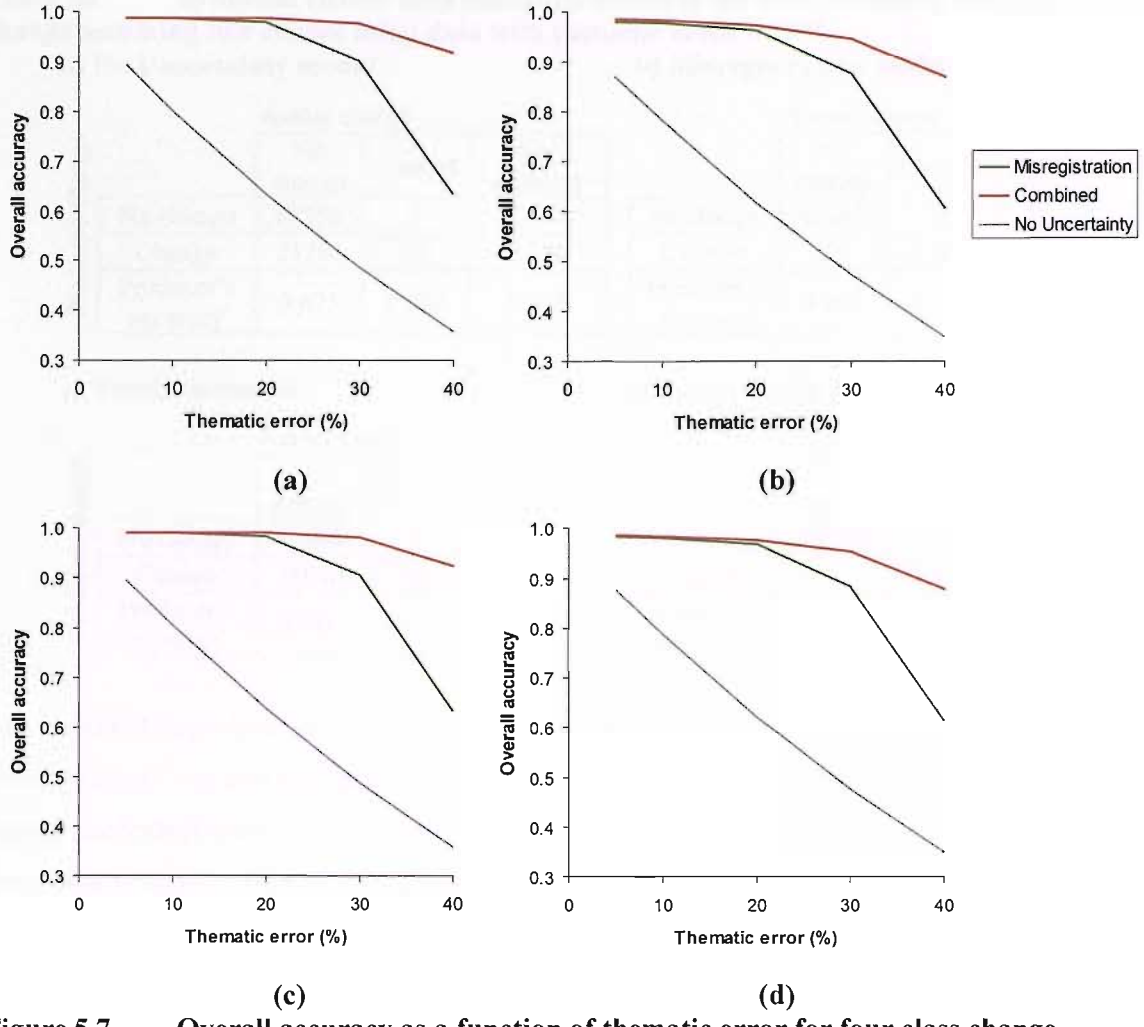


Figure 5.7 Overall accuracy as a function of thematic error for four class change model.
a) Time 1, hard boundary.
b) Time 2, hard boundary. Boundary between classes moved 40 m.
c) Time 1, fuzzy boundary width = 20 m.
d) Time 2, fuzzy boundary width = 20 m. Boundary between classes moved 40 m.

Using the Misregistration method when the thematic error was small was as accurate as or more accurate than using the Combined method (Figure 5.6; Figure 5.7). However, as thematic error increased, the Combined method became more accurate than the Misregistration method alone, indicating that the effect of thematic uncertainty in change detection increased with increasing thematic error (Figure 5.6). In the four-class problem this effect was increased and the Misregistration method became much less accurate relative to the Combined method at large thematic errors (Figure 5.7).

Table 5.1 Synthetic change data confusion matrices for hard boundary and no change assuming two classes using data with thematic error = 20%.

a) No Uncertainty model

Predicted change \ Actual change	No change	Change	User's accuracy
No change	44252	0	1.000
Change	21284	0	0.000
Producer's accuracy	0.675	0.000	0.675

b) Misregistration model

Predicted change \ Actual change	No change	Change	User's accuracy
No change	65460	0	1.000
Change	76	0	0.000
Producer's accuracy	0.999	0.000	0.999

c) Thematic model

Predicted change \ Actual change	No change	Change	User's accuracy
No change	46906	0	1.000
Change	18630	0	0.000
Producer's accuracy	0.716	0.000	0.716

d) Combined model

Predicted change \ Actual change	No change	Change	User's accuracy
No change	65531	0	1.000
Change	5	0	0.000
Producer's accuracy	1.000	0.000	1.000

The increased importance of using thematic uncertainty measures in change detection as thematic error increased was also observed when comparing the No Uncertainty and Thematic change methods (Figure 5.6). At small thematic errors, the two methods had similar accuracies for change detection. However, as thematic error increased, the difference in change detection accuracy increased, with the Thematic uncertainty method having a larger accuracy than the No Uncertainty method.

The reduction in accuracy as a function of an increase in thematic error also occurred in the prediction of the thematic change vector magnitude (Figure 5.8). All methods of change detection became less accurate as the thematic error increased.

The relative importance of thematic uncertainty as thematic error increased was also observed in the plots of the RMSE between predicted and actual thematic change vectors, particularly the plots of thematic change vector error as a function of thematic error for the areas of change only (Figure 5.9). At small thematic errors the Misregistration method resulted in smaller errors in the predicted thematic change vector than the Combined method, but as thematic error increased, the Misregistration method became less accurate compared to the Combined method (Figure 5.9).

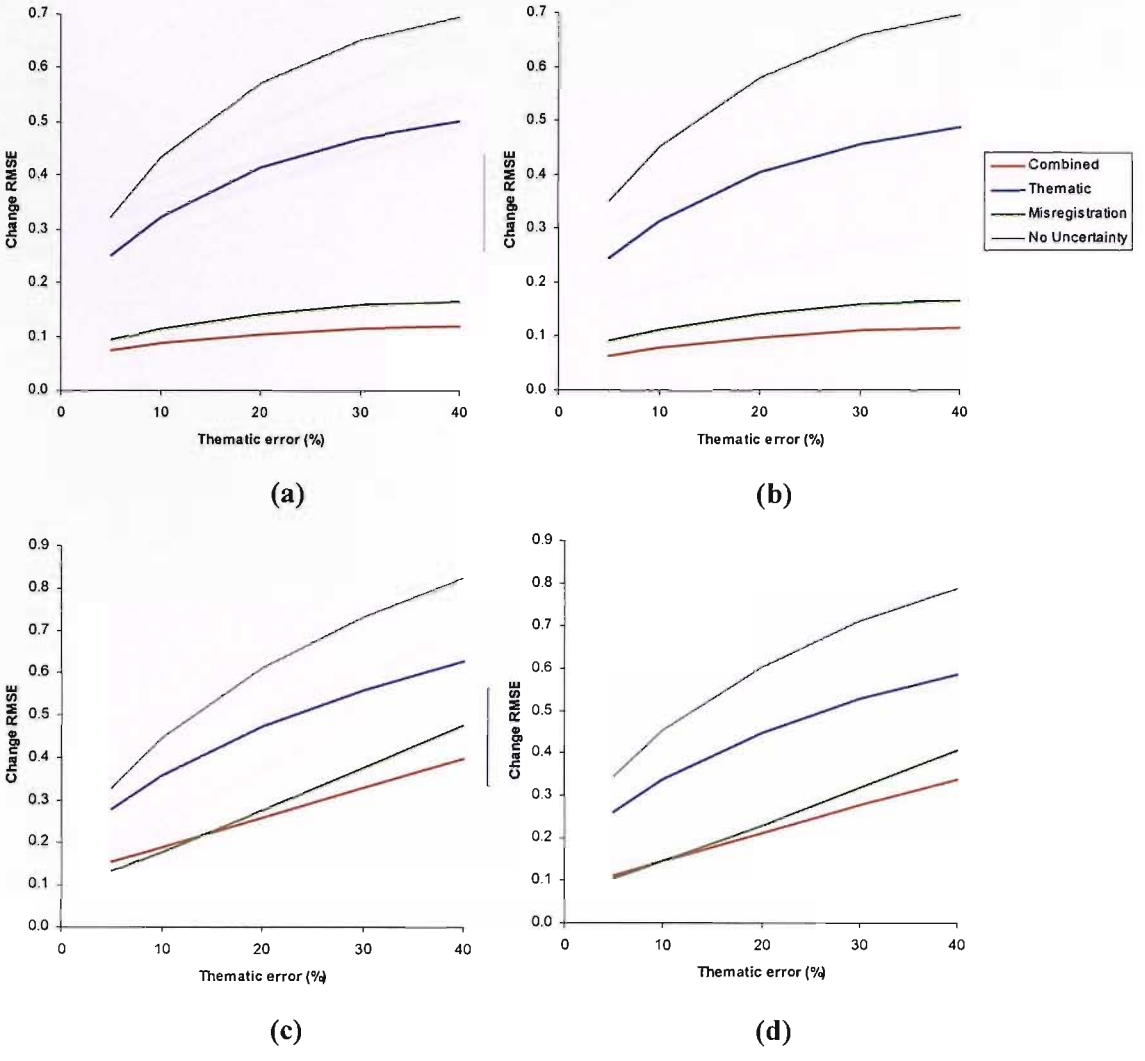


Figure 5.8 RMSE as a function of thematic error.
a) Time 1, hard boundary. Boundary between classes moved 5 m.
b) Time 2, fuzzy boundary width = 20 m. Boundary between classes moved 5 m.
c) Time 1, hard boundary. Boundary between classes moved 40 m.
d) Time 2, fuzzy boundary width = 20 m. Boundary between classes moved 40 m.

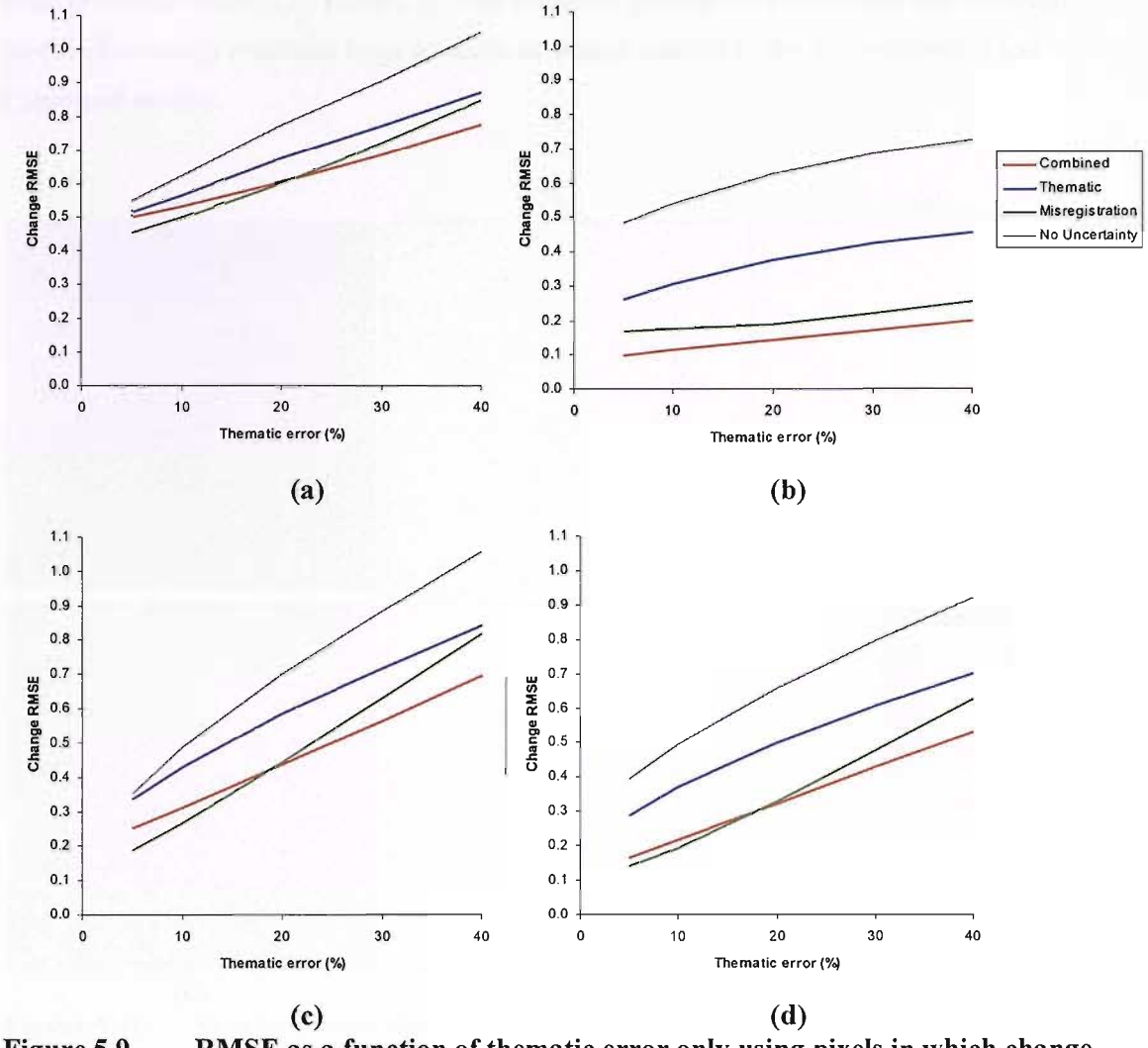


Figure 5.9 RMSE as a function of thematic error only using pixels in which change took place.
a) Time 1, hard boundary. Boundary between classes moved 5 m.
b) Time 2, fuzzy boundary width = 20 m. Boundary between classes moved 5 m.
c) Time 1, hard boundary. Boundary between classes moved 40 m.
d) Time 2, fuzzy boundary width = 20 m. Boundary between classes moved 40 m.

5.6.2 Impact of fuzziness and change

5.6.2.1 Hard boundaries with no thematic change

In the images where no change took place and the classes had a hard boundary, change was incorrectly predicted at points across the whole image for the No Uncertainty and Thematic models (Figure 5.10). This also occurred in the transect data, with large negative and positive predicted thematic change vectors in areas of no actual change (Figure 5.11). The confusion matrices for these models show that the Producer's accuracy for the No Change class was much greater using the Misregistration and Combined methods classes in both the two-class and four-

class problems (Table 5.1; Table 5.2). This indicated that the No Uncertainty and Thematic models incorrectly predicted large amounts of change relative to the Misregistration and Combined models.

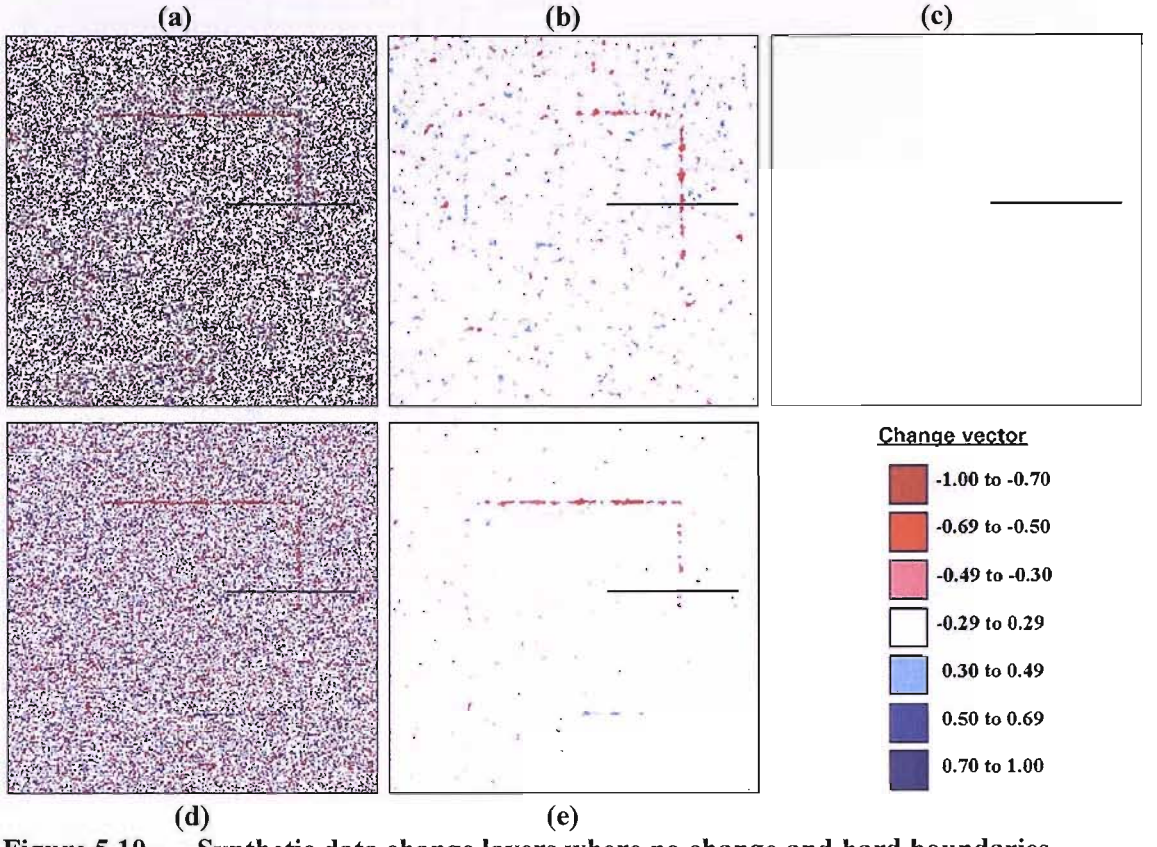


Figure 5.10 Synthetic data change layers where no change and hard boundaries between classes. Black lines are positions of transects in Figure 5.11.
a) No Uncertainty model.
b) Misregistration model.
c) Actual change.
d) Thematic model.
e) Combined model.

The Misregistration change model reduced the incorrect prediction of change compared to the No Uncertainty and Thematic models, with the Combined method reducing these errors further (Figure 5.10). These trends also occurred in the transect data (Figure 5.11), with the variation from the actual thematic change vector being least in the Combined model and only slightly greater in the Misregistration model. The No Uncertainty model had the greatest variations from the actual thematic change vector with errors that were slightly greater than the Thematic model (Figure 5.11). The observations above match the confusion matrices of the different change detection methods (Table 5.1; Table 5.2).

Table 5.2 Synthetic change data confusion matrices for hard boundaries between classes and no change assuming four classes using data with thematic error = 20%.

a) No Uncertainty and Thematic models.

Predicted change \ Actual change	No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	14006	0	0	1815	0.885
Change 1-2	3486	0	0	7052	0.000
Change 2-1	3535	0	0	7211	0.000
No change 2-2	877	0	0	27554	0.969
Producer's accuracy	0.639	0.000	0.000	0.632	0.634

b) Misregistration model

Predicted change \ Actual change	No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	21438	0	0	61	0.997
Change 1-2	208	0	0	221	0.000
Change 2-1	238	0	0	536	0.000
No change 2-2	20	0	0	42814	1.000
Producer's accuracy	0.979	0.000	0.000	0.981	0.980

c) Combined model

Predicted change \ Actual change	No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	21538	0	0	30	0.999
Change 1-2	102	0	0	117	0.000
Change 2-1	248	0	0	307	0.000
No change 2-2	16	0	0	43178	1.000
Producer's accuracy	0.983	0.000	0.000	0.990	0.987

Areas incorrectly predicted as change occurred at the boundary between the two classes using all change detection methods (Figure 5.10). A large decrease in incorrect prediction of change occurred when the Misregistration and Combined models of change detection were used compared to the Thematic and No Uncertainty models of change detection (Figure 5.10; Figure 5.11). Though the Combined model removed most of the incorrect prediction of change in the homogeneous areas, it still incorrectly predicted change at the boundaries, but the predicted thematic change vector was generally less than 0.5 (Figure 5.10). This error may be seen in the transect data (Figure 5.11). At the boundary between the classes (50 m), a large negative thematic change vector occurred in the Misregistration and Combined change methods, showing that at the boundaries these methods did not completely remove the effects of misregistration (Figure 5.11).

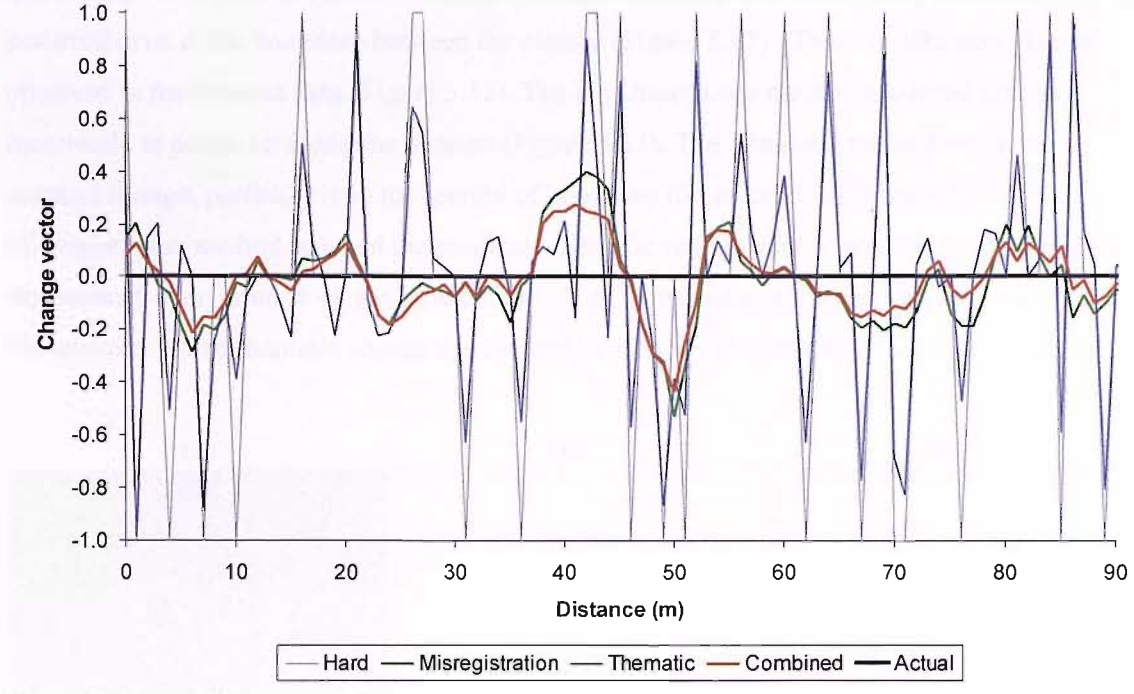


Figure 5.11 Change vector along transects for No change and hard boundaries between classes. Boundary at 50 m. Transect position shown in Figure 5.10. 0 m is on the left side of the transect.

Very little change was predicted incorrectly using the Combined or Misregistration methods for the two-class and four-class problems, with Producer's and User's accuracies for the No change class all greater than 97.8% (Table 5.1; Table 5.2). This compared with Producer's accuracy of 71.6% for the Thematic method and 67.5% for the No Uncertainty method using the two-class problem. In the four-class problem using the No Uncertainty method, the Producer's accuracy of the No change classes were 63.9% and 63.2%. These results showed that when using the No Uncertainty and Thematic methods a large number of pixels were misallocated to the Change classes.

5.6.2.2 Fuzzy boundaries with no thematic change

In the images where no change occurred and the classes had a 20 m fuzzy boundary width (Figure 5.12) change was only predicted incorrectly at the boundaries using the No Uncertainty model. Using the Misregistration model a small increase in incorrect change prediction occurred at one part of the boundary on the right side of the image (Figure 5.12). A small reduction in the incorrect change prediction occurred over the fuzzy boundary using the Thematic model, which may be observed as a faint square in the middle of the image (Figure 5.12). This was likely to occur due to the thematic uncertainty values being closer to one another for t_1 and t_2, due to fuzziness in pixels and a corresponding reduction in predicted change in the thematic

uncertainty. Using the Combined method when the boundary was fuzzy, very little error occurred even at the boundary between the classes (Figure 5.12). These results may also be observed in the transect data (Figure 5.13). The No Uncertainty method predicted change incorrectly at points all along the transect (Figure 5.13). The Thematic method predicted reduced change, particularly in the section of boundary (35 m to 55 m) (Figure 5.13). The Misregistration method reduced the predicted thematic change vector below 0.5 along most of the transect, apart from at 48 m (Figure 5.13). The Combined method reduced the error all along the transect, with a thematic change vector smaller than 0.4 (Figure 5.13).

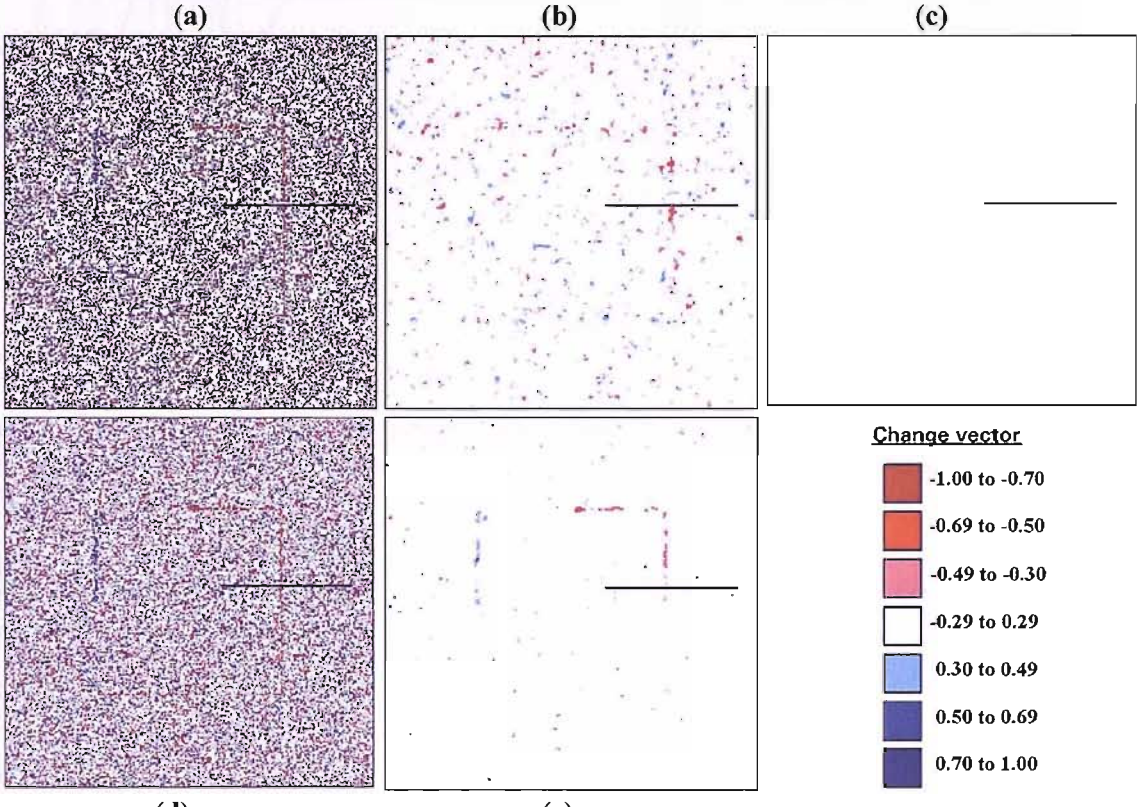


Figure 5.12 Synthetic data change layers where there is no change and fuzzy boundary width = 20 m. Black lines are positions of transects in Figure 5.13.

a) No Uncertainty model.
b) Misregistration model.
c) Actual change.
d) Thematic model.
e) Combined model.

In the images with no actual change the Misregistration and Combined methods removed much of the error due to change being incorrectly predicted (Figure 5.10; Figure 5.12). This was also observed in the confusion matrices for the two and four-class change problems when the thematic error was 20% and the classes had a hard boundary (Table 5.3; Table 5.4). In the two-class and four-class change problems, the results of the no change scenario for a thematic error

of 20% where the classes had fuzzy boundaries were similar to the results where the classes had hard boundaries (Table 5.1; Table 5.2; Table 5.3; Table 5.4).

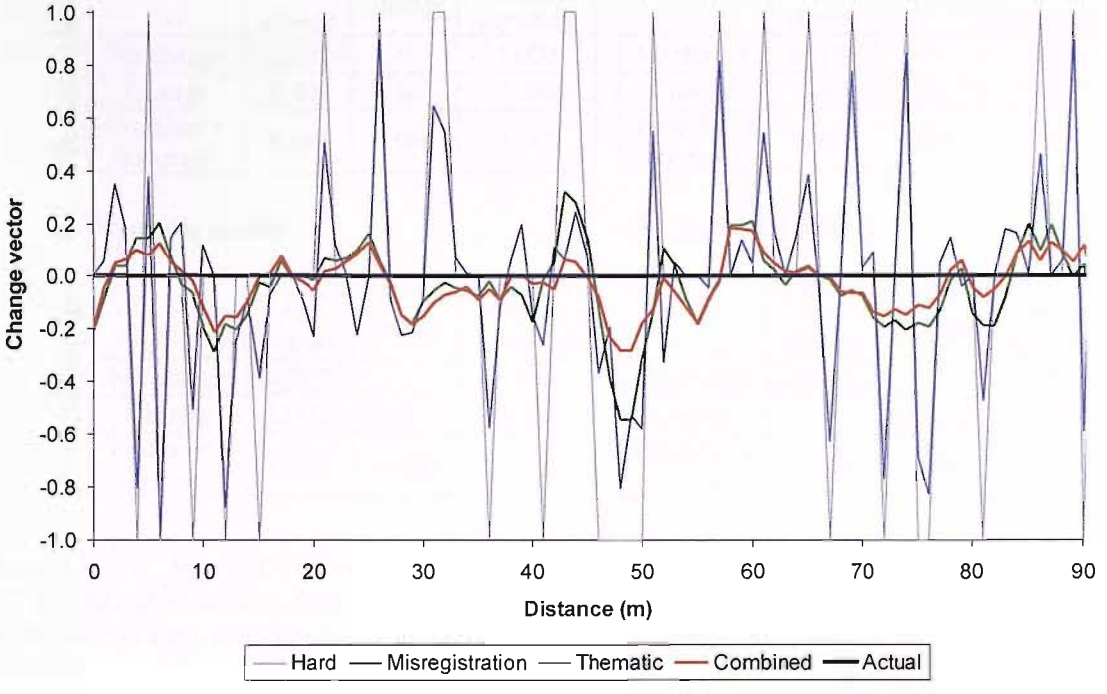


Figure 5.13 Change vector along transects for No change and fuzzy boundary width = 20 m. Boundary between 35 m and 55 m. Transect position shown in Figure 5.12. 0 m is on the left side of the transect.

Using the Combined or Misregistration methods the Producer's and User's accuracy for the No change class was greater than 99% for the two-class problem and 92.8% for the four-class problem (Table 5.3; Table 5.4). This compared with a Producer's accuracy of 73.6% for the Thematic method and 66.7% for the No Uncertainty method using the two-class problem (Table 5.3). In the four-class problem using the No Uncertainty change method when there were fuzzy boundaries, for the No Change classes the Producer's accuracies were 59.6% and 62.4% and the User's accuracies were 81.9% and 96.7% (Table 5.4). These results were similar to those where classes had hard boundaries (Table 5.2) showing that unlike the Combined and Misregistration methods, using No Uncertainty and Thematic methods a large number of pixels were misallocated to the Change classes.

Table 5.3 Synthetic change data confusion matrices for fuzzy boundary width = 20 m and no change assuming two classes using data with thematic error = 20%.

a) No Uncertainty model

Predicted change	Actual change: No change	Actual change: Change	User's accuracy
No change	43697	0	1.000
Change	21839	0	0.000
Producer's accuracy	0.667	0.000	0.667

b) Misregistration model

	Actual change: No change	Actual change: Change	User's accuracy
No change	65490	0	1.000
Change	46	0	0.000
Producer's accuracy	0.999	0.000	0.999

c) Thematic model

Predicted change	Actual change: No change	Actual change: Change	User's accuracy
No change	48225	0	1.000
Change	17311	0	0.000
Producer's accuracy	0.736	0.000	0.736

d) Combined model

	Actual change: No change	Actual change: Change	User's accuracy
No change	65536	0	1.000
Change	0	0	0.000
Producer's accuracy	1.000	0.000	1.000

Table 5.4 Synthetic change data confusion matrices for fuzzy boundary width = 20 m and no change assuming four classes using data with thematic error = 20%.

a) No Uncertainty and Thematic models.

Predicted change	Actual change: No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	10079	0	0	2222	0.819
Change 1-2	2835	0	0	7953	0.000
Change 2-1	2960	0	0	8091	0.000
No change 2-2	1026	0	0	30370	0.967
Producer's accuracy	0.596	0.000	0.000	0.624	0.617

b) Misregistration model

	Actual change: No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	15683	0	0	153	0.990
Change 1-2	404	0	0	337	0.000
Change 2-1	557	0	0	710	0.000
No change 2-2	256	0	0	47436	0.995
Producer's accuracy	0.928	0.000	0.000	0.975	0.963

c) Combined model

Predicted change	Actual change: No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	15889	0	0	131	0.992
Change 1-2	317	0	0	227	0.000
Change 2-1	494	0	0	446	0.000
No change 2-2	200	0	0	47832	0.996
Producer's accuracy	0.940	0.000	0.000	0.983	0.972

5.6.2.3 Hard boundaries with thematic change

In the change images where 40 m of change took place and the classes had hard boundaries, the general pattern of actual change was observable in all images (Figure 5.14). In the homogeneous areas of no change, similar patterns occurred to those in Figure 5.10 and Figure 5.12. In the area of actual change, the No Uncertainty and Thematic methods of change detection predicted a large proportion of pixels as undergoing no change (Figure 5.14). For many of the pixels, the predicted thematic change vector was opposite to the actual thematic change vector (Figure 5.14). The predicted change was Class 2 to Class 1, when the actual change was Class 1 to Class 2.

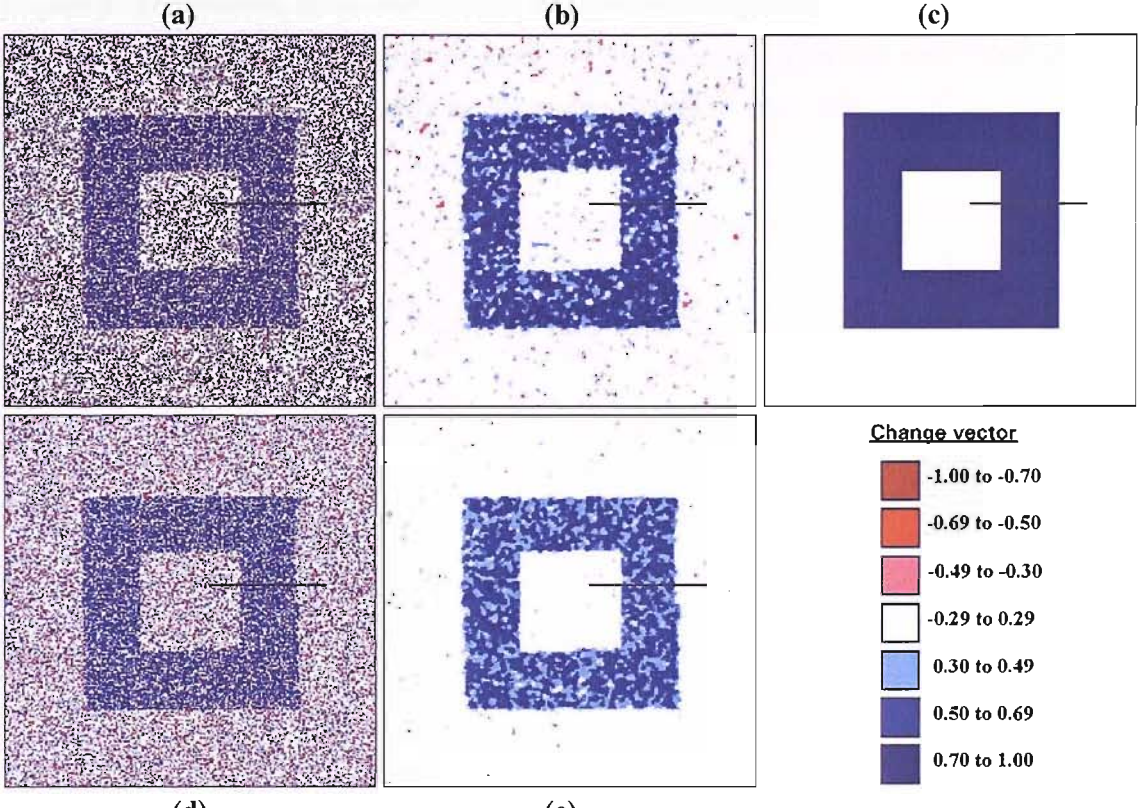


Figure 5.14 Synthetic data change layers where change = 40 m and hard boundaries between classes. Black lines are positions of transects in Figure 5.15.
a) No Uncertainty model.
b) Misregistration model.
c) Actual change.
d) Thematic model.
e) Combined model.

For both the Misregistration and Combined change detection methods the predicted change was less than the actual change, though the thematic change vectors generally had the correct sign for much of the change area (Figure 5.14). This indicated that the No Uncertainty model was

less accurate than the other methods at predicting the direction of the thematic change vector (which classes a pixel was allocated at t_1 and t_2).

The exception was the edge of the area of change, where change was predicted less accurately using the Misregistration and Combined methods. This may be observed in the transect data (Figure 5.15). At the 60 m point on the transect the Misregistration and Combined methods did not predict change to the edge of the change area (Figure 5.15). At the points where the No Uncertainty and Thematic methods correctly predicted change, these methods had a thematic change vector closer to the actual value than either the Misregistration or Combined methods (Figure 5.15). However, the overall error of the thematic change vector for the Misregistration and Combined methods was smaller than the No Uncertainty and Thematic uncertainty methods as may be seen for the RMSE values for the area of change (Figure 5.9).

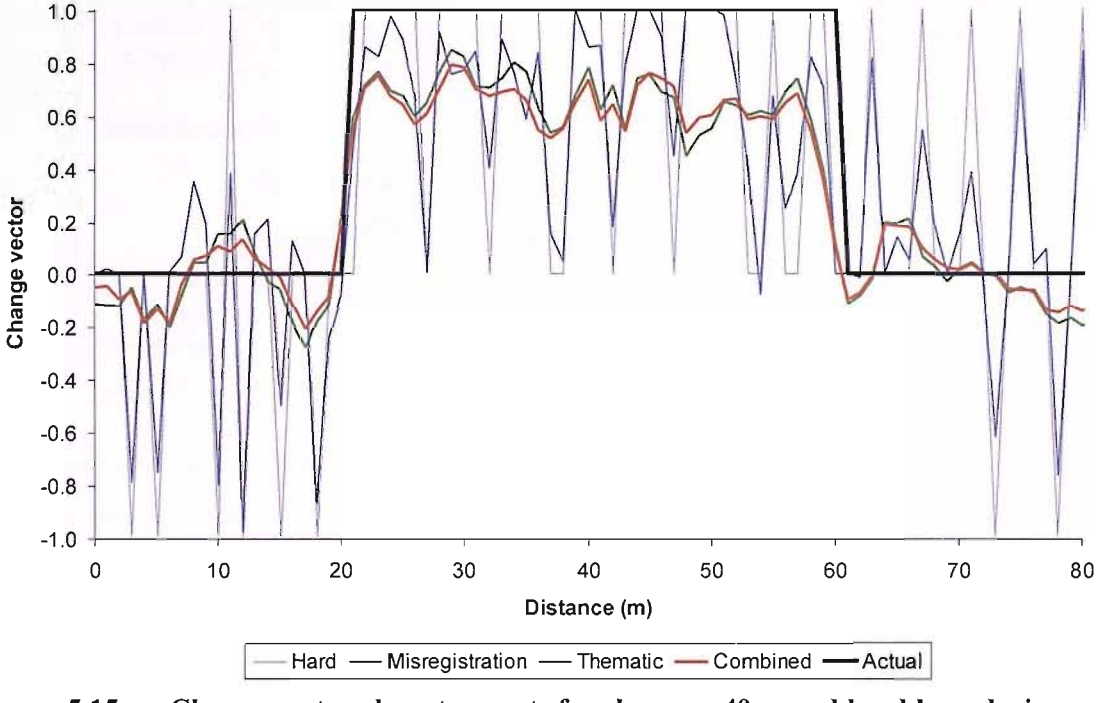


Figure 5.15 Change vector along transects for change = 40 m and hard boundaries between classes. Transect position shown in Figure 5.14. 0 m is on the left side of the transect.

When change occurred for the classes with the hard boundaries, the No Uncertainty and Thematic change methods had large errors. In the two-class problem the errors were largest in the Change class with Producer's accuracies of 65.4% and 67.8% and User's accuracies of 45.1% and 42.9% for the Thematic and No Uncertainty methods respectively (Table 5.5). The majority of errors in the Misregistration and Combined change methods were predicting areas of

change as undergoing no change (Table 5.5). These methods removed most incorrect predictions of change, with the Change class User's accuracies of 99.8% and 99.5% for the Combined and Misregistration methods respectively (Table 5.5). The No change User's accuracies were still large, with values of 92.8% (Combined) and 91.2% (Misregistration). However, the Producer's accuracies for the Change class were much smaller, with values of 78.2% (Combined) and 73% (Misregistration) indicating that these methods underestimated change (Table 5.5).

Table 5.5 Synthetic change data confusion matrices for classes with hard boundary and change = 40 m assuming two classes using data with thematic error = 20%.

a) No Uncertainty model

Predicted change	Actual change: No change	Change	User's accuracy
No change	32630	5562	0.854
Change	15626	11718	0.429
Producer's accuracy	0.676	0.678	0.677

b) Misregistration model

	Actual change: No change	Change	User's accuracy
No change	48198	4673	0.912
Change	58	12607	0.995
Producer's accuracy	0.999	0.730	0.928

c) Thematic model

Predicted change	Actual change: No change	Change	User's accuracy
No change	34524	5981	0.852
Change	13732	11299	0.451
Producer's accuracy	0.715	0.654	0.699

d) Combined model

	Actual change: No change	Change	User's accuracy
No change	48230	3764	0.928
Change	26	13516	0.998
Producer's accuracy	0.999	0.782	0.942

The accuracy of change detection when using the four-class problem was greater than the two-class problem for both the Misregistration (92.8% and 98.3%) and Combined (94.2% and 99%) methods (Table 5.6). For both these change methods, the Change 1-2 class had the smallest Producer's accuracy (Table 5.6). However, unlike the two-class problem (Table 5.5), the amount of misclassification of areas of Change as areas of No change was more evenly matched, with 457 Change pixels misclassified as No change and 213 No change pixels misclassified as areas of Change in the Combined model (Table 5.6). Using the Misregistration model 541 Change pixels were misclassified as No change and 561 No change pixels were misclassified as areas of Change (Table 5.6).

Table 5.6 Synthetic change data confusion matrices for classes with hard boundary and change = 40 m assuming four classes using data with thematic error = 20%.

a) No Uncertainty and Thematic models.

Predicted change \ Actual change	No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	2972	2775	0	1784	0.395
Change 1-2	764	10981	0	7083	0.583
Change 2-1	718	737	0	7061	0.000
No change 2-2	170	2787	0	27704	0.904
Producer's accuracy	0.643	0.635	0.000	0.635	0.636

b) Misregistration model

Predicted change \ Actual change	No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	4551	285	0	4	0.940
Change 1-2	72	16738	0	279	0.979
Change 2-1	1	1	0	209	0.000
No change 2-2	0	256	0	43140	0.994
Producer's accuracy	0.984	0.969	0.000	0.989	0.983

c) Combined model

Predicted change \ Actual change	No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	4575	193	0	0	0.960
Change 1-2	49	16823	0	147	0.988
Change 2-1	0	0	0	17	0.000
No change 2-2	0	264	0	43468	0.994
Producer's accuracy	0.989	0.974	0.000	0.996	0.990

5.6.2.4 Fuzzy boundaries with thematic change

In the change images where 40 m of change took place and the fuzzy boundary width was 20 m, the general pattern of actual change occurred again in all images (Figure 5.16). In the homogeneous areas of no change, the same patterns occurred as in all the previous synthetic change diagrams (Figure 5.10; Figure 5.12; Figure 5.14). The No Uncertainty and Thematic methods of change detection predicted a large proportion of Change pixels as being No change or predicted a thematic change vector with the opposite sign to the actual thematic change vector (Figure 5.16). As would be expected, the No Uncertainty change method did not show the fuzzy nature of change and below the 0.5 threshold generally did not predict any change as having occurred. The Thematic method did show the fuzzy nature of the change, as the thematic change vectors at the edge of the edge of the change area were more representative of the actual thematic change vectors than using the No Uncertainty change method (Figure 5.16).

The Misregistration change method did not predict the fuzzy nature of the actual change as accurately as the Combined method (Figure 5.17). Though thematic change vectors at the edge of the change area had a smaller magnitude than the change vectors in the centre of the change

area, these did not indicate the fuzzy nature of the change (Figure 5.17). The Combined method appeared to represent the fuzzy nature of the change at the edge of the area of change more accurately than the Misregistration method (Figure 5.16; Figure 5.17). However, as with the change when there were hard boundaries (Figure 5.14), the Combined and Misregistration methods consistently underestimated the magnitude of the thematic change vector (Figure 5.16). This may be seen in the transect data, where for most of the change area actual thematic change vectors were greater than those predicted (Figure 5.17).

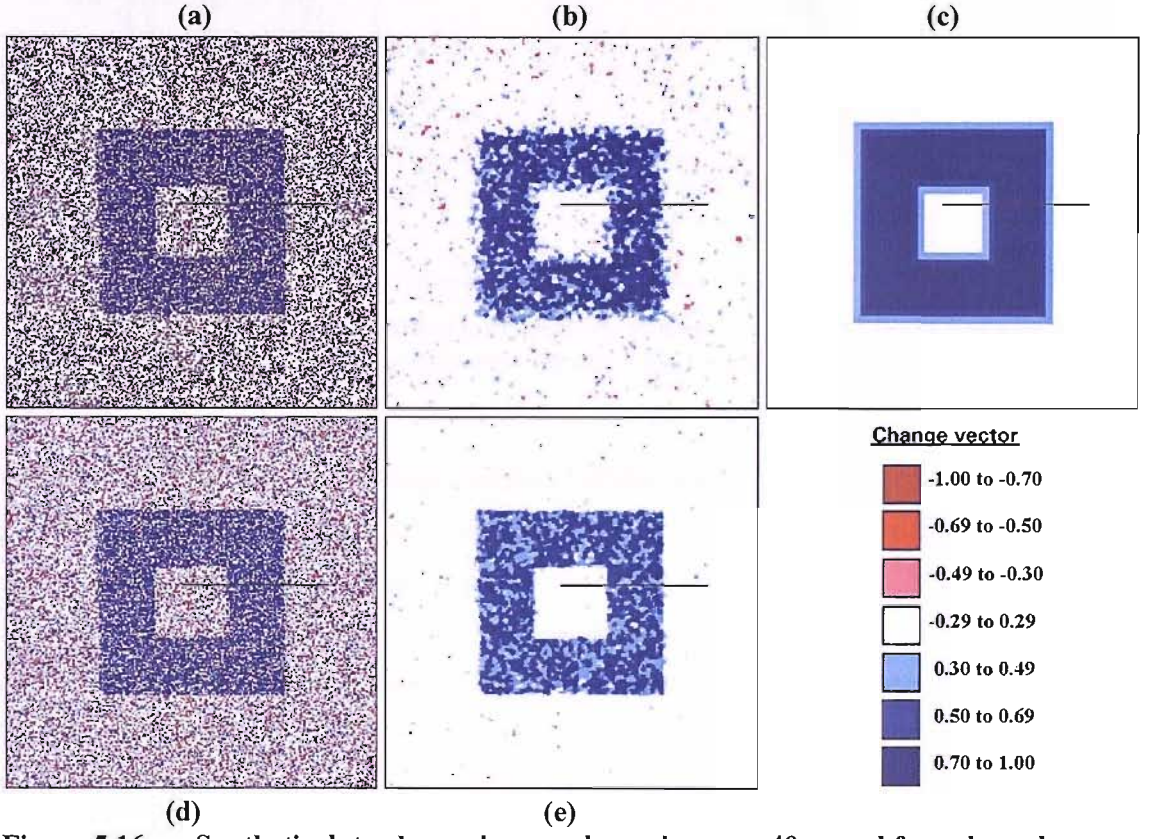


Figure 5.16 Synthetic data change layers where change = 40 m and fuzzy boundary width = 20 m. Black lines are positions of transects in Figure 5.17.

a) No Uncertainty model.
b) Misregistration model.
c) Actual change.
d) Thematic model.
e) Combined model.

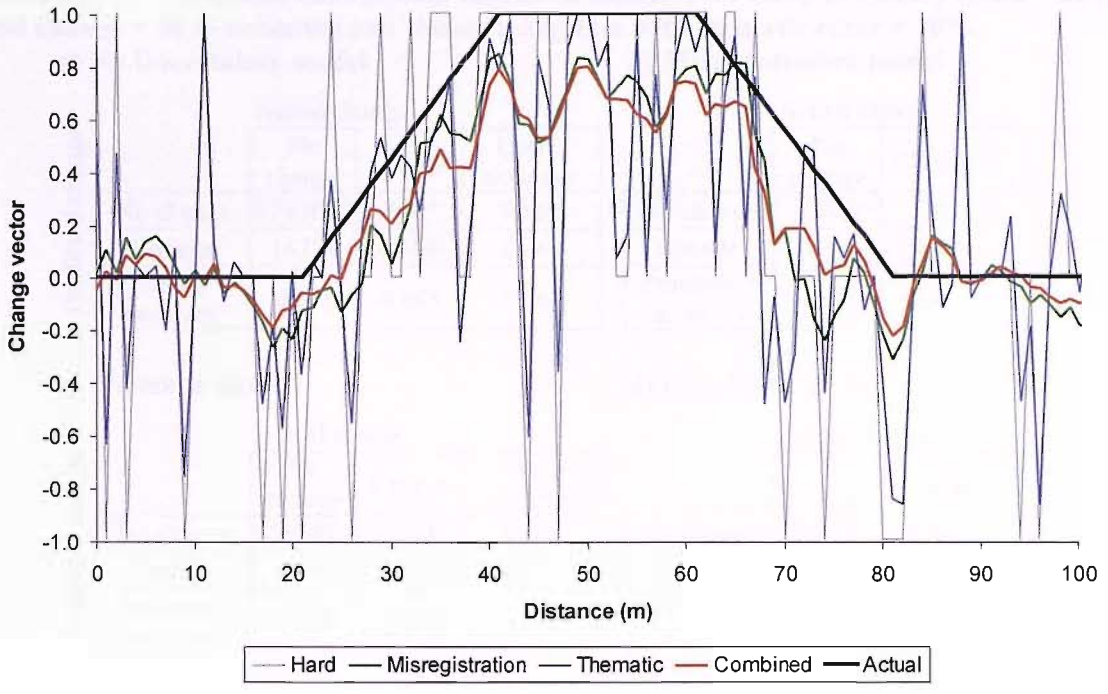


Figure 5.17 Change vector along transects for change = 40 m and fuzzy boundary width = 20 m. Transect position shown in Figure 5.16. 0 m is on the left side of the transect.

The No Uncertainty model of change detection had the largest Producer's accuracy for the Change class (64.5%) in the two-class problem (Table 5.7). This model detected more of the actual change than any of the other models when the boundary was fuzzy. However, when looking at the four-class problem using the No Uncertainty model, the Producer's accuracy for the class Change 1–2 was much smaller (59.7%) than the value obtained when the Combined (93.9%) and Misregistration (92.5%) models were used (Table 5.8). Though the No Uncertainty model could detect much of the change, it did not predict what change occurred. This was supported by the prediction using the No Uncertainty model of 8857 pixels of class Change 2-1, compared to 231 pixels using the Misregistration model and 17 pixels using the Combined model, when in the actual data no pixels of this class occurred (Table 5.8). Though the No Uncertainty model predicted more of the actual change in the two class problem than the other models, the User's accuracy of the Change class in both the No Uncertainty (36.0%) and Thematic models (37.3%) was very small compared to the Misregistration (98.9%) and Combined models (100%) (Table 5.7). In the two-class problem the actual change was 22.3% of the total area. The No Uncertainty and Thematic models of change predicted 39.9% and 34.9% change respectively, compared to the 12.4% by the Combined model and 14.2% by the Misregistration model (Table 5.7).

Table 5.7 Synthetic change data confusion matrices for fuzzy boundary width = 20 m and change = 40 m assuming two classes using data with thematic error = 20%.

a) No Uncertainty model

Predicted change \ Actual change	No change	Change	User's accuracy
No change	34201	5187	0.868
Change	16739	9409	0.360
Producer's accuracy	0.671	0.645	0.665

b) Misregistration model

Predicted change \ Actual change	No change	Change	User's accuracy
No change	50842	5387	0.904
Change	98	9209	0.989
Producer's accuracy	0.998	0.631	0.916

c) Thematic model

Predicted change \ Actual change	No change	Change	User's accuracy
No change	36593	6074	0.858
Change	14347	8522	0.373
Producer's accuracy	0.718	0.584	0.688

d) Combined model

Predicted change \ Actual change	No change	Change	User's accuracy
No change	50937	6501	0.887
Change	3	8095	1.000
Producer's accuracy	1.000	0.555	0.901

Table 5.8 Synthetic change data confusion matrices for fuzzy boundary width = 20 m and change = 40 m assuming four classes using data with thematic error = 20%.

a) No Uncertainty and Thematic models.

Predicted change \ Actual change	No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	1439	2317	0	2042	0.248
Change 1-2	566	8592	0	8133	0.497
Change 2-1	377	719	0	7761	0.000
No change 2-2	118	2772	0	30700	0.914
Producer's accuracy	0.576	0.597	0.000	0.631	0.622

b) Misregistration model

Predicted change \ Actual change	No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	2233	275	0	1	0.890
Change 1-2	266	13313	0	489	0.946
Change 2-1	1	7	0	223	0.000
No change 2-2	0	805	0	47923	0.983
Producer's accuracy	0.893	0.925	0.000	0.985	0.968

c) Combined model

Predicted change \ Actual change	No change 1-1	Change 1-2	Change 2-1	No change 2-2	User's accuracy
No change 1-1	2245	182	0	0	0.925
Change 1-2	255	13524	0	358	0.957
Change 2-1	0	0	0	17	0.000
No change 2-2	0	694	0	48261	0.986
Producer's accuracy	0.898	0.939	0.000	0.992	0.977

5.6.3 Global measures of change

The effects of the different models in terms of global change detection were summarised by comparing the amount of change predicted by each of the change models with the actual amount of change for the classifications with 20% thematic error (Table 5.9; Table 5.10). In both the two-class and four-class problems, the Combined and the Misregistration approaches only slightly over-predicted change when no change took place (Table 5.9; Table 5.10). The Thematic model overestimated change in all cases, and the No Uncertainty model overestimated change more than all other models (Table 5.9; Table 5.10).

When change did take place the Thematic and No Uncertainty models overestimated the total area of change. In the two-class problem the Combined and Misregistration models both underestimated change (Table 5.9). The Combined model (20.7%) was closest to the actual change (26.4%) when the boundaries were hard, and the Misregistration model (14.2%) was closest to the actual change (26.4%) when the boundaries were fuzzy (Table 5.9). When the results of the four-class problem were examined, both the Combined and Misregistration models were much closer to the actual change than with the two-class problem (Table 5.10). In both the hard and fuzzy boundary change scenarios the Misregistration model was within 0.2% of the actual value and Combined model within 0.4% (Table 5.10). The total percentage change was more accurately modelled using the four-class rather than the two-class approach.

Table 5.9 Percentage change using synthetic change data two-class problem, thematic error = 20%

Actual Change (m)	Fuzzy boundary width (m)	Combined (%)	Thematic (%)	Misregistration (%)	No Uncertainty (%)	Actual change (%)
0	0	0.0	28.4	28.4	0.1	32.5
0	20	0.0	26.4	26.4	0.1	33.3
40	0	20.7	38.2	38.2	19.3	41.7
40	20	12.4	34.9	34.9	14.2	39.9

Table 5.10 Percentage change using synthetic change data four-class problem, thematic error = 20%

Actual Change (m)	Fuzzy boundary width (m)	Combined (%)	Misregistration (%)	Thematic/No Uncertainty (%)	Actual change (%)
0	0	1.2	1.8	32.5	0.0
0	20	2.3	3.1	33.3	0.0
40	0	26.0	26.4	41.7	26.4
40	20	21.6	21.8	39.9	22.0

5.7 Discussion and Conclusions

The use of synthetic data enabled variables to be controlled that would be difficult or impossible to control using environmental data. However, assumptions were made during the process of generating the synthetic data that may have resulted in outputs that were not representative of actual airborne data. Though synthetic data have the potential to provide information about the change detection models derived in this study, care should be taken in any conclusions drawn.

As thematic error increased, the ability to detect change decreased, even when thematic and geometric uncertainty measures were used (Figure 5.6; Figure 5.7). This indicates that even when using thematic and geometric uncertainty measures, the maximum precision of change detection is a function of the accuracy of the input data. Though methodologies were developed to minimise the impact of errors, these were not able to remove the effect of the error in the input data.

The impact of using thematic uncertainty varied as a function of thematic error. When comparing overall accuracy and RMSE as a function of thematic error (Figure 5.6; Figure 5.7; Figure 5.8; Figure 5.9) the Misregistration model was as accurate as or more accurate than the Combined model at small thematic errors. However, as thematic error increased, the Combined model became more accurate relative to the Misregistration method. The importance of using thematic uncertainty as thematic error increased was also observed when the No Uncertainty and Thematic change methods were compared (Figure 5.6; Figure 5.7). When thematic errors were small, the two methods had similar levels of accuracy. As thematic error increased, the difference in accuracy increased, with the Thematic uncertainty method having a larger accuracy than the No Uncertainty method. This indicates that the suitability of using thematic uncertainty varies according to thematic accuracy. As thematic error increased, the use of thematic uncertainty when modelling change was likely to result in more accurate modelling of change.

For all accuracy measures used, the change detection model that used combined per-pixel thematic and geometric uncertainty was more accurate than the change detection models that assumed no uncertainty or only used thematic per-pixel uncertainty. One of the reasons for this was likely to be that the use of misregistration uncertainty smoothed thematic errors. By applying the misregistration function, the thematic errors were smoothed across a wide area, reducing their per-pixel effect. This may be seen in the change images and the change transects (Figure 5.10 to Figure 5.17). In the simple synthetic data model with spatially random thematic

error, this reduced change detection error, but in environmental data it is likely to reduce the precision of the model, as areas of any class with a width of a similar order to the misregistration error are likely to be smoothed such that they will be removed. However, in a pixel based method of detecting change there will always be compromise between reduction of misregistration errors (generally incorrect detection of change at boundaries) and missing small, pixel order changes.

The Thematic and No Uncertainty change methods generally overestimated change (Table 5.9; Table 5.10). This was likely to be because errors generally resulted in the incorrect prediction of change. For a given pixel where no actual change took place, a thematic error in only one of the input classifications would result in a 100% probability of change being predicted. For a pixel where actual change took place, an error in one of the input classifications would only result in a prediction of no change if the pixel was allocated the same class as the other input layer. The probability of this happening will be inversely proportional to the number of classes, at largest 50% for two classes. As a majority of thematic errors will appear as change, this is likely to bias outputs resulting in overestimation of change.

The Misregistration and Combined models generally underestimated change. This is because these models smoothed the change surface, making them less sensitive to differences between the input layers. Class boundaries were expanded beyond their actual position to compensate for potential geometric error. This will result in ecotones appearing wider than they actually are and result in a loss of sensitivity at boundaries particularly where change has taken place. The edges of areas of actual change are less likely to be detected than the centre. This was observed in the images of change and the change transects (Figure 5.14 to Figure 5.17), where change became less obvious at the edge of the change areas, particularly where change occurred over a wide rather than narrow ecotone.

The misregistration uncertainty model smoothed errors due to incorrect detection of change. This may be seen in the plot of accuracy as a function of change detection accuracy for the two-class problem and no change (Figure 5.6). Even as thematic error was increased, the accuracy of change detection did not decrease (Figure 5.6). This smoothing effect may result in removal of small objects with a width of the order of a few pixels, therefore increasing thematic error. If the geometric error footprint can be reduced, the smoothing effect will be reduced increasing the precision of change detection.

5.7.1 Summary

In this chapter measures of misregistration and thematic uncertainty were tested for change detection using synthetic data. Change detection with and without misregistration and thematic uncertainty measures was tested. By adding misregistration and thematic uncertainty measures in a change detection model the accuracy of change detection could be increased.