UNIVERSITY OF SOUTHAMPTON

# REGIONALISING A DAILY RAINFALL RUNOFF MODEL WITHIN THE UNITED KINGDOM

by

**Andrew Richard Young**

A thesis submitted for
the degree of
Doctor of Philosophy

Department of Geography

October 2000

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE
GEOGRAPHY

Doctor of Philosophy

# REGIONALISING A DAILY RAINFALL RUNOFF MODEL WITHIN THE UNITED KINGDOM

by Andrew Richard Young

Access to daily stream flow data, at the river reach scale is a central component of many aspects of water resource and water quality management. However, the majority of river reaches within the UK are ungauged. Hence, there is an operational requirement for a quick, consistent and reliable method for simulating historical stream flow records within ungauged catchments. The overall objective of this research has been to develop a rainfall runoff model for predicting natural daily stream flows within a catchment without recourse to the calibration of model parameters against observed stream flow data. Implicit within this objective is a requirement that the model parameters can be estimated from readily available data describing the physical characteristics of the catchment.

The fundamental approach to the research has been to develop and calibrate suitable models within a large, representative sample of UK catchments and to subsequently develop predictive, statistical relationships for estimating model parameters from the climatic and physiographic characteristics of the catchments. The predictive capacity of the regionalised model forms has been extensively evaluated through comparisons with gauged flow data, calibrated models and existing industry-standard methods for estimating historical flow time series within ungauged catchments.

The regionalised model forms developed represent a significant advance over existing, low-cost methods for estimating historical flow regimes within ungauged river catchments. The errors in the simulated stream flows are sufficiently small for the techniques to be a useful aid in the management of water resources within the UK.

# Contents

# Figures

# Tables

# Acknowledgements

I would like to thank my Ph.D. supervisor, Nigel Arnell for his support. I would also like to thank Alan Gustard, CEH - Wallingford, not only for his support and advice in respect of this research, but also for his support and friendship over the last 10 years at CEH.

The ideas and focus of the research within this thesis have been developed over a 10-year period of working with staff from the Environment Agency for England and Wales. The original ideas were first discussed during a project, titled "Naturalised River Flow Records of the Essex Region", commissioned by Nigel Fawthrop in the early 1990s. A collaborative research programme, jointly funded by CEH and the Environment Agency, in the area of low river flows provided the means to undertake the bulk of the research within this thesis. I would like to thank the CEH-Agency Low Flows Steering Committee for their comments and suggestions over the last four years. In this context I would particularly like to thank Dr. Glenn Watts and Dr. Rob Grew.

Thanks are given to my employers, CEH-Wallingford for sponsoring this part time Ph.D. There are many friends at CEH that I would like to thank. I would particularly like to thank Matt Holmes and Ann Sekulin for their encouragement and Matt for his willingness to read and comment on early drafts of the thesis.

These acknowledgements would be incomplete without thanking Janet for her love and support of this project, which over the last year must have appeared as a petty distraction during some very grim times. I would like to show my gratitude to the boys, Zac and Sam, for accepting that Dad has had to work on "the thesis" instead of going to the park. Finally, I would like to dedicate this thesis to our daughter Anna and the joy that she brings to us all.

# List of Abbreviations and Definitions

## ABBREVIATIONS

| | |
|---|---|
| AAR | Average Annual Rainfall |
| ALTBAR | Mean catchment altitude |
| AREA | Catchment Area |
| ARRO | Average Annual Runoff |
| ASPBAR | Mean direction of all 50m slopes within a catchment |
| ASPVAR | The invariability of catchment slope directions |
| BEQ95 | BIAS error between simulated and observed Q95 flows |
| BFI | Base Flow Index |
| BFIHOST | Estimate of BFI derived from HOST |
| BIAS | error between simulated and observed mean flows expressed as a percentage of the observed mean flow |
| CV | Coefficient of Variation |
| DBFI | The difference between observed and simulated BFI |
| DPLBAR | The mean of the drainage network distances measured between each node of a 50m grid |
| DPLCV | The coefficient of variation of the drainage network distances measured between each node of a 50m grid |
| DPSBAR | The mean of all the inter-nodal slopes for the catchment |
| EFF | The Nash-Sutcliffe Efficiency Criterion |
| FDC | Flow Duration Curve |
| HOST | Hydrology of Soil Types Classification |
| HOST | Hydrology of Soil Types |
| HOSTRES | the residual from a linear regression model fitted between SPRHOST and BFIHOST for 687 catchments |
| HYSIM | Hydrological Simulation Model |
| IH-ACRES | Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Stream flow data |
| LDP | Longest Drainage Path within a catchment |

| | |
|---|---|
| LF_OBJ | The mean sum of squared differences between observed and simulated flows for observed flows that are less than the flow that is equalled or exceeded for 67% of the time. |
| MODA | five parameter model developed for regionalisation |
| MODB | three parameter model developed for regionalisation |
| MORECS | Meteorological Office Rainfall and Evaporation Calculation System |
| MORECS | Meteorological Office Rainfall and Evaporation Calculation System |
| P_AREA | The difference between the catchment areas for two catchments expressed as a percentage of the smallest value of catchment area for the pair. |
| P_RUNOFF | the difference between runoff estimates for two catchments expressed as a percentage of the smallest value of catchment runoff for the pair. |
| PDM | Probability Distributed Model |
| PDSM | Probability Distributed Soil-moisture Model |
| PP | Potential for Soil Moisture Deficits |
| Q95 | The flow that is equalled or exceeded for 95% of the time |
| Q95%MF | Q95 flow normalised by the mean flow |
| SAAPE(61-90) | Standard Period Average Annual Potential Evaporation over the period 1961-1990 |
| SAAR(61-90) | Standard Period Average Annual Rainfall over the period 1961-1990 |
| SMD | Soil Moisture Deficit |
| SPR | Standard Percentage Runoff |
| SPRHOST | Estimate of SPR derived from HOST |
| TCM | Thames Catchment Model |
| VAR | normalised non-parameterically derived measure of variance |

DEFINITIONS FOR MODA AND MODB MODEL PARAMETERS

| Parameter name | Unit | Description |
|---|---|---|
| *Interception store* | | |
| $\gamma$ | mm | Depth of interception store in MODA |
| *Probability-distributed soil Moisture store* | | |
| Cmax | mm | The maximum store capacity within the catchment. |
| b | none | The exponent of the Pareto distribution, controlling the spatial variability of store capacities. |
| kg | hour | The groundwater recharge time constant for MODA. |
| $\beta$ | none | The runoff factor, which controls the split of direct runoff between surface and groundwater storage routing for MODB. |
| *Routing Module* | | |
| K1 | hour | The time constant for the quick flow linear reservoir. |
| Kb | hour | The time constant for the slow flow, or base flow linear reservoir |

# 1 Introduction

## 1.1 OVERVIEW OF RESEARCH OBJECTIVES

Information on the magnitude and variability of flow regimes, at the river reach scale is a central component of most aspects of water resource and water quality management. For some activities, such as the setting of discharge consents and licencing of small abstractions, it is sufficient to encapsulate this information using a statistical description of the flow regime. The flow duration curve (NERC, 1980) is an example of this type of analysis. However, there are many applications for which a time series of stream flows is required. These include the assessment of yield for water resource schemes, the in-stream flow requirements of aquatic flora and fauna and the assessment of the impacts of climate change at the catchment scale.

At the broadest scale, natural river flow regimes are dependent on rainfall, temperature and evaporation. At the catchment scale, the flows will be controlled by the physical properties of a catchment, including geology and land use. A rainfall runoff model predicts stream flow by using a mathematical description of the processes controlling the catchment stream flow response to climatic inputs of precipitation and evaporation demand. The response of the model is controlled by parameters, which normally have to be estimated through calibration.

The overall research objective of this thesis has been to develop a rainfall runoff model for predicting the natural variation of daily stream flows within a catchment without recourse to the calibration of model parameters against observed stream flow data. Implicit within this objective is a requirement that the model parameters can be estimated from the physical characteristics of the catchment. The focus of the research has been on an operational requirement for a quick, consistent and reliable method for simulating historical stream flow records within ungauged catchments. Hence, as part of this overall objective, it is essential that the data required by the model are restricted to those data that are readily available in a digital form.

For water management purposes, it is essential to differentiate between the natural and artificial components of stream flow data. The artificial component is the nett influence of water use within the catchment. Common influences include surface and groundwater abstractions, discharges from sewage treatment plants and industrial sources, impounding reservoirs, canal transfers and inter-basin transfer schemes. A review of over 1600 gauging stations has identified that less than 20% of gauged catchments within the UK can be regarded as being natural (Gustard *et al.*, 1992). In assessing the available resource within a catchment, it common practice to separate the natural and artificial components of stream flow. This separation of components enables practitioners to assess the natural reliable yield of the catchment, based upon the climatically driven variability of the natural stream flow. The impacts of actual, and planned water use scenarios are subsequently superimposed upon the natural flow regime to assess both the reliable yield under the current water use and/or scenarios of future water use and the environmental impacts of that water use.

The temporal resolution for both managing and regulating water use is commonly between a week and a month. For assessing the yield of some storage schemes, it is not necessary to consider stream flow at a finer temporal resolution than this; for example the estimation of natural inflows into an impounding reservoir. However, for run-of- river water use schemes, and cases where the impact of a scheme on aquatic flora and fauna is to be assessed it is essential to consider stream flow at a finer temporal resolution. The commonly used resolution is a calendar day. This finer resolution is important for assessing the frequency of failure of a scheme (for example a direct abstraction), the ecological impacts of flow derogation and, in the case of discharges, the impact of flow on water quality and hence the flora and fauna of the stream.

In estimating the time series of natural stream flows at a site it is not necessary to exactly replicate all aspects of the true natural time series. Rather, the requirement is to simulate important facets of the regime including:

- an acceptable simulation of mean flow– conservation of mass;
- how the stream flow reduces in the absence of rainfall- termed recession behaviour;
- the correct representation of seasonal patterns within the flow regime;
- the correct stream flow response to precipitation and the dependencies of that response on antecedent catchment conditions.

With regard to the last point, it is not important to accurately simulate individual high flow events. The only restrictions on the modelling of high flows are that mass must be conserved over a longer time period and the observed sequencing of high flow events should be replicated. In the context of run-of-river schemes, the high flows are not a resource that can be readily utilised, due to the high concentrations of suspended solids. As the cause-effect links between flow and habitat for aquatic species cannot be accurately quantified, predictive methods for assessing the ecological impacts of high flows are not sensitive to the absolute magnitude of the flows.

The industry best current practice for estimating the natural stream flow record from a gauged record is to naturalise the flow record through decomposition (Hall and Nott, 1994; Young and Sekulin, 1996). This technique is based on identifying the nett influence within the catchment above a gauging station and subtracting that nett influence from the gauged flow record. As data describing the volumetric impact of artificial influences are commonly inaccurate and archived at a monthly resolution, the technique is both time consuming and error prone.

Within the UK there are approximately 1600 permanent gauging stations, these gauged river reaches represent less than one percent of the total number of river reaches mapped at a scale of 1:50,000. The majority of water management decisions are therefore being taken within catchments for which there are no measured stream flow data. Historically, practitioners have quantified flow regimes within ungauged catchments using simple, statistically based models for predicting natural and artificially influenced flow statistics

(Natural Environment Council, 1980, Southern Water Authority, 1979; Gustard and Sutcliffe 1986; Gustard *et al* 1987, 1992; Holmes and Young, 2000a&b). Currently, the low cost solution to the requirement for historical time series of river flows at ungauged catchments is to transpose gauged data from a similar, nearby gauged catchment. Commonly, these gauged flows will be naturalised prior to transposition.

The objective of this research has been to develop a model for predicting stream flows within ungauged catchments directly from historical climatic data and the physical characteristics of the catchment. The science of relating hydrological phenomena to physical and climatic characteristics of a catchment, or region, is commonly called regionalisation. As will be discussed, there are many examples within the international literature of studies directed at the regionalisation of rainfall runoff models. These have generally been undertaken at a spatial or temporal scale that is inappropriate for catchment scale water management.

Experience has shown that, for a study into the regionalisation of a daily conceptual rainfall runoff model to be successful and substantive, the following aspects must be addressed.

- The model must be applied to a representative sample of catchment types across the country. To date regionalisation studies have been restricted to relatively small sample sets of catchments.
- In catchment specific applications, the model must be able to accurately simulate those aspects of the flow regime that are of importance for water resource assessment and management decisions – and this must hold over the full range of catchment types.
- The calibrated parameter vectors must, within the uncertainty of the input data, be identifiable for specific catchment types, and the model fit must be stable when applied to an independent evaluation period.
- The model structure, and hence the parameters must have physical meaning to enable the parameters to be estimated from the physical characteristics that can be used to differentiate between catchment types.

The scope of the research has been restricted to catchments in which precipitation falls primarily as rainfall, rather than as snow. This restriction is only an issue for some highland catchments within the north of the country. The research has addressed the estimation of climatic data with the emphasis on selecting/deriving methods for estimating climatic data using modifications of best practice techniques that do not incur a prohibitive cost. The solution used for estimating evaporation demand has meant that catchments within Northern Ireland have been excluded from the research as historical MORECS (see Chapter 2) data are not available for Northern Ireland. The evaluation of success of the research is based on whether:

- the regionalised model forms developed represent an advance over existing methods for estimating historical flow regimes within ungauged river catchments;
- The errors in the simulated stream flows are sufficiently small for the techniques to be a useful aid in the management of water resources within the UK.

An overview of rainfall runoff modelling and the issues associated with the structure and calibration of this class of model are discussed in Section 1.2. Previous regionalisation studies within the UK, and elsewhere within the world are reviewed within Section 1.3. The innovative aspects of this thesis are discussed within Section 1.4. The structure of the thesis is also presented within this section.

## 1.2 RAINFALL RUNOFF MODELLING

The objective of rainfall runoff modelling, in it's broadest sense, is to simulate the translation of precipitation, that is incident upon the surface of a stream or river catchment, to stream flow at the catchment outlet taking into account evaporative losses from the system. Todini (1988) gives a useful review of the development of the science of rainfall runoff modelling and modelling philosophies. He introduces the concept of a mathematical model consisting of two parts; one part consisting of the physical model structure which encapsulates the *a priori* knowledge of the system and the second part a stochastic component that cannot be explained by the physical model structure. This led him into a four-class model classification scheme based on the degree of prior knowledge. This classification is summarised in Figure 1.1. The system differentiates between models on

the initial basis of whether processes are represented statistically or physically and then how these processes are distributed and solved mathematically across the catchment. Other classifications have been proposed by Singh (1995) and Hughes (1995), amongst others.

In reality, the boundaries between classifications and between the individual boxes within a classification are not clearly defined. Many models are essentially hybrid with constituent parts drawing from stochastic and deterministic components. The deterministic components may seek to describe the physics of the process, using differential equations, commonly called physically based, or may use a conceptual representation of the physical processes, in which integral equations are used to represent the processes. Physically based models are distributed in that the model equations include space co-ordinates and are differential in nature (thus requiring the definition of boundary conditions) as opposed to integral, as in the case of the conceptual rainfall runoff model. Additionally, stochastic techniques are now commonly used when formulating the catchment implementation of deterministic model components; for example, the semi-distributed soil moisture module of the Probability Distributed Model (Moore, 1985) and the ARNO (Todini, 1996), XINANJIANG (Zhao, 1980) and VIC (Wood, 1992) models.

Singh in his 1995 paper states:

"*A vast majority of the (available) models are deterministic, and virtually no model is fully stochastic. In some cases, only some parts of the model are described by the laws of probability, and other parts are fully deterministic. It is then fair to characterise them as quasi-deterministic or quasi-stochastic.*"

In summary, it is true to say that models grade in the complexity (both with respect to the model structure and spatial resolution) from purely empirical statistical models through to differential physically based distributed models. Conceptual models, whether lumped or with some degree of spatial discretisation, lie between these extremes. In the context of this study a model is considered as "lumped" if the input data, output data and model equations do not include a spatial description. This definition does not make a distinction between stochastic or deterministic formulations.

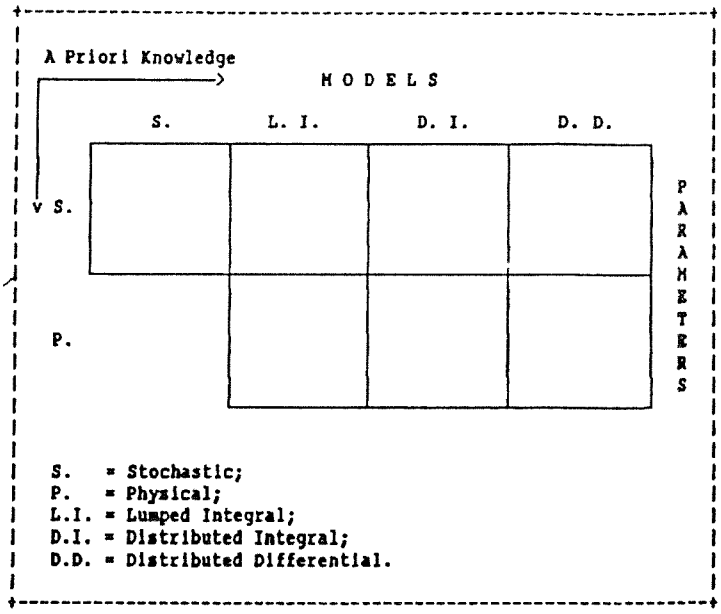



**Figure 1.1 Classification based on level of prior knowledge** ***(Source: Todini, 1988).***

An argument can be constructed that no model components are truly physically based. Any mathematical description of a process is an approximation of that process and thus is always a conceptualisation. The preservation of the physicality of physically based deterministic model components is called into question in the application of the model. Whilst the process descriptions may model the transport of water under well-defined laboratory conditions, they may not when applied to the complexities of a real catchment. The scale of the spatial and temporal discretisation of the model is extremely important. In practice, it is necessary to limit the resolution of distributed models to a grid scale that is commensurate with the input data describing catchment properties, the climatological variations and computing power available. This lumping and the uncertainty in the input climatic data and field measurement of catchment properties (and hence parameter values) will generally mean that the model will require calibration to compensate for these uncertainties (Beven, 1993). Hence, the true physicality of the model is compromised. For a further discussion of the relative merits of this type of model compared with the simpler, integral conceptual class of deterministic models the reader is directed to the work of Beven (1989 & 1993), Todini (1988) and Singh (1995) as an inroad to this topic. This study is concerned with the question of the regionalisation of lumped, conceptual models on a scale where the application of differential distributed models is not a practical proposition.

## 1.3 REGIONALISATION OF RAINFALL RUNOFF MODELS

The development of a model of the rainfall runoff process that can be applied without recourse to calibration data, and hence applied within un-gauged catchments, has been the subject of research since the 1960s. The theory is that if the structural description is correct the parameters of the model are more likely to be related to physical characteristics of the catchment that can be measured. This has lead in many cases to large, complex models that, from a systems-engineering point of view, are over parameterised. By a systems engineering point of view it is meant that the parameters of the model are calibrated in accordance with some scheme to minimise the differences between observed and simulated stream flows. As discussed in the previous section a degree of calibration is necessary with all models.

The international literature on the regionalisation of rainfall runoff models falls into two basic categories.

- The calibration of a model on a range of catchments types and the subsequent development of statistical relationships between model parameters and the physical and climatological characteristics of the catchments. These relationship then enable model parameters to be estimated for the un-gauged site.

- The *a priori* estimation of model parameters, that purport to have physical realism, through direct measurement of the physical characteristics of the catchment to which the parameters pertain.

Models have been regionalised both for simulating catchment response to extreme precipitation events at an hourly time step and for simulating the much longer-term temporal variation of river flows, at either a daily or longer time step, for resource evaluation purposes. These studies have all generally been conducted using a relatively small number of study catchments (less than 40). The studies that are of relevance to the current study, irrespective of the purpose, are reviewed here according to purpose.

### 1.3.1 Model regionalisation for flood estimation.

The United Kingdom Flood Studies Report (FSR) (Natural Environment Research Council, 1975) presents a method for flood frequency analysis based on the use of a unit hydrograph model to estimate the peak flow corresponding to a design rainfall event. The two primary parameters for this model are the "Time to Peak" parameter for the unit hydrograph, Tp and the standard percentage runoff, SPR. A third parameter, the peak flow of the one-hour unit hydrograph, Qp is estimated from Tp. Within the FSR, SPR and Tp are estimated from regression relationships for five geographic regions within the UK. These regression relationships were revised in the Flood Studies Supplementary Report No. 16 (Institute of Hydrology, 1985).

Burn and Boorman (1993) used the regionalisation of Tp and SPR as a vehicle for evaluating methods, other than regression relationships, for classifying catchments and estimating model parameters. In this approach, 99 catchments within the UK were clustered into groups based upon their hydrological similarity, as represented by principal components of key catchment characteristics. A derivative of discriminant analysis was then used to assign catchments to the predefined groups. Several options were then used, based on the knowledge of group membership and the characteristics of the groups, to estimate the model parameters for catchments that were treated as ungauged catchments. The baseline for evaluation of various options was the revised FSR regression models for predicting Tp and SPR. The study concluded that SPR was estimated most efficiently by identifying the groups of nearest neighbour catchments and taking a weighted (based on distance in catchment characteristic space) average of the SPR parameters for those catchments. The results for estimating Tp were more ambiguous with no particular option being identified as the preferred approach. The study did not consider the impact of the various parameter estimation methods on the simulation of stream flow for example events.

Pirt and Bramley (1985) present regression-based equations for estimating parameters for an eight-parameter version of the isolated event model. These equations were derived through applying the model to fourteen small sub-catchments within the River Trent. The relationships were evaluated by applying the regionalised model over three events within two further catchments within the Trent basin and two catchments from Yorkshire. The

catchment characteristics used within the study were based on data describing the variations in topography, catchment geometry and soils. The evaluation gave good results in two of the catchments and poor results in the other two catchments, and must therefore be regarded as being inconclusive.

Hughes (1989) adopted a similar approach to that of Pirt and Bramley for an eight-parameter isolated event model called OSE2. In this study, the model was applied to 29 catchments from a study set of 21 catchments drawn from the United States (Vermont, Arizona, Mississippi and Oklahoma) and 12 catchments from South Africa. Relationships between model parameters and catchment characteristics were then derived graphically. These relationships were based on readily derived deterministic measures, such as drainage density, and empirical "scores" or indices of characteristics, such as soil depth and channel roughness. Hughes draws a comparison between the results of his study and those of Pirt and Bramley, stating that, whilst his results were substantially poorer, his study covered a much greater physiographic and climate range in terms of the catchments considered.

Calver *et al* (1999) presents the results from a pilot study of the regionalisation of two models for the purposes of flood frequency estimation. The models were the TATE model (Calver, 1996) and a version of the PDM model of Moore (1985). The study considered 40 catchments and 379 station-years of continuous hourly data. Model parameters were related, using multivariate regression relationships, to the Flood Estimation Handbook catchment descriptors discussed in Chapter 5 of this thesis. The performance of the regionalised models was assessed by the capacity of the models for accurately predicting flood frequency distributions.

### 1.3.2 Model regionalisation for annual, monthly or daily flow regime estimation

Jarboe and Hann (1974) report the results of study in which a four-parameter water yield model, running on a monthly time step, was regionalised using a group of 24 catchments in the State of Kentucky. An approach of calibrating the model on 17 of the catchments and subsequently relating the model parameters to catchment characteristics using multivariate regression models was adopted. The regionalised model was evaluated against the gauged annual runoff within the remaining catchments. Errors were in the range 2-12 percent. The

authors highlight the model deficiencies and rainfall errors as being the limiting factor on the results and acknowledge the limited sample size of the study in restricting the wider application of the approach. The study presented results for annual water balance predictions only.

Magette *et al* (1976) present a regionalisation of the Kentucky Watershed Model (a derivative of the Stanford Watershed model of Crawford and Linsley (1966)) based on 21 catchments. The regionalised model was used for estimating mean annual flows (aggregated from the model when run on a daily time step) and stream flow response to specific storm events (using an hourly time step). The calibrated model parameters were statistically related to physical and land use characteristics of the catchments. Five catchments were retained as an independent test data set. The results obtained were inconclusive, there were appreciable errors in predicted mean annual runoff (>20%) for four catchments and a mixture of reasonable and poor simulations of storm events. Egbuniwe and Todd (1976) present the calibration of the Stanford Watershed Model IV on two Nigerian catchments that have similar highly seasonal climatic regimes but dissimilar hydrogeological controls. The authors transposed the parameters from one catchment to the other and evaluated the model fit obtained with the transposed parameters. An acceptable water balance was obtained, but there were appreciable discrepancies between the observed and simulated monthly and daily time series.

In the UK Manley (1977) presents the application of the HYSIM model (See Chapter 3) to the River Dove, which is an 883 $km^2$ tributary of the River Trent. Manley partitioned the basin into a number of gauged sub-catchments and used an *a priori* definition of most parameters using the channel, soils and geological characteristics of the Dove and the gauged flows from the adjacent River Derwent catchment. He concluded that that over half of the *a priori* parameters were within a factor of 1.5 of the calibrated values and that both calibrated and estimated parameter gave acceptable simulations.

The approach reported by Tulu (1991) is similar to that of Egbuniwe and Todd (1976). The author calibrated an existing deterministic conceptual model against monthly stream flow data for the Guder River (central Ethiopian Highlands) and then applied the model using the Guder parameter set to a tributary of the basin, the Teltele basin, with satisfactory

results. There will obviously be a high degree of serial correlation between the stream flows within the two basins due to their nested nature.

Agung and Cordery (1995) describe the regionalisation of a simple four parameter lumped conceptual model run on a monthly time step. The model was calibrated on 18 catchments with New South Wales (Australia) covering catchment areas between 10 and 1530 $km^2$. In South Africa, Hughes and Sami (1995) describe the conceptual basis and structure of a semi-distributed, conceptual model, VTI-HYMAS that can be run on a variable time step. In the paper the authors discuss the physical relevance of the model parameters and outline a procedure for parameter estimation. The Assessment of Surface Water Resources of South Africa (Midgley *et al*, 1994) describes a regionalisation approach for a version of the monthly time step Pitman model (Pitman and Kakebeeke, 1991). In this approach, the drainage basins in South Africa were sub-divided in to Quaternary catchments. These represent the fourth, and finest, sub-division of catchments for the assessment. The model was calibrated for these catchments. The catchments were subsequently empirically grouped into hydrologically similar groups based on climate, topography, soils, geology and vegetation. Model parameters were then averaged within groups for application at ungauged sites.

Post and Jakeman (1996) present graphical relationships obtained between the calibrated model parameters for a second order configuration of the IH-ACRES model (described within Chapter 3) and physical catchment characteristics for sixteen small hydrogeologically homogenous catchments within the state of Victoria, Australia. The physical catchment characteristics used described the drainage network, catchment geometry and vegetation classes within the catchments. Sefton and Boorman (1997) present the results from a study into the feasibility of regionalising IHACRES within the UK. The study presents the results from applying a second order configuration of IHACRES to 39 natural catchments which have catchment areas of less than 1000$km^2$. The model parameters were related to catchment characteristics using linear regression modelling. Within the paper the regression relationships are graded from good to poor but the utility of the relationships in predicting stream flow in ungauged catchments was not directly evaluated.

### 1.3.3 Model regionalisation for linking with Global Circulation Models.

One class of model in which research is aligned to the objectives of this study is the macro-scale model. Arnell (1999) provides a useful review of this type of model. He defines a macro-scale model as a model that can be applied repeatedly over a large geographic domain, without the need for local calibration. Macro-scale models are regionalised models of land-surface hydrological processes that can simulate these processes accurately at an appropriate temporal resolution, and that can be incorporated within climate simulation models. The model maybe run at a daily time step, however the model output is aggregated into a monthly or annual time series for subsequent use. The end point use of such models is primarily, but not exclusively, for predicting the effect of future climate change. This necessitates the use of a deterministically based representation of the hydrological processes involved in the rainfall-runoff process. This class of model is generally based on a gridded representation of the modelling of soil moisture availability. Each grid cell is treated as a discrete entity, although cell outputs may subsequently be passed though a routing mechanism (Jolley and Wheater, 1997a&b).

Arnell (1999) points out that the availability of input data will significantly constrain the form of a macro-scale model, and that this will commonly determines the grid scale. Jolley and Wheater (1997a) investigated the effect of spatial scale on the performance of a 1-D water balance model based upon the Penman drying curve concept (Penman, 1949). This concept has been widely used within the Thames Catchment Model (TCM) (Greenfield, 1988) and is discussed in more detail within Chapter 3. The study indicated that the performance of the model was more sensitive to the averaging of the climatic data in higher rainfall catchments and to the averaging of soil moisture behaviour in lower rainfall catchments.

The use of a distribution function to describe the spatial variation in grid based storage capacities has been successfully used by Arnell in his derivatives of Moore's Probability Distributed Model (Moore, 1985). This work was initially undertaken in the UK (Arnell and Reynard, 1996) and has since been developed to give global coverage (Arnell and King, 1997). Wood has also successfully applied the approach in the VIC model (Wood *et al* 1992). In application, Arnell assigns *a priori* parameter estimates that are either fixed or

allowed to vary through space. These parameter estimates are derived from catchment scale experience of applying the PDM and process based catchment studies. Abdulla *et al* (1997a &b) have demonstrated for a 2 layer version of the VIC model how this *a priori* approach may be expanded in continental scale river basins where good quality data describing the climatic variations and physical characteristics of catchments are available. Within the model, a Pareto distribution function is used to describe the variability of storage capacities across a grid cell of specified resolution. The study was based upon application of a lumped version of the model to 34 natural catchments within the Arkansas-Red River basin within the United States. The model was run on a daily time step with the outputs aggregated to monthly values for calibration against observed monthly stream flow data. Of the nine model parameters, two were defined *a priori* from catchment soils information and the remainder calibrated against stream flow data. The parameters were then subsequently related to catchment characteristics using multivariate regression techniques. The authors make the point that climate can play an important role in model descriptions of hydrological response, primarily due to limitations of models in describing the full complexities of the rainfall runoff process.

Abdulla (1997b) presents an assessment of the performance of the Macro version of VIC-2L within the Red River basin. The assessment covered the case when the model was run with parameter fields derived from the regression relationships and the case when the model was run with parameter fields derived by interpolation from the calibrated catchments. The comparison demonstrated that the use of the regression models gave acceptable simulations of monthly hydrographs and significantly improved performance in reducing model bias.

There is a conceptual problem that needs to be considered when grid based macro models are run with parameters that have been estimated on the basis of the behaviour of the model when applied in a lumped mode. The problem is that the model structure is conceptually different when applied within a uniform grid mode. In the lumped mode the parameters are a function of the whole catchment processes. If the grid resolution is small compared to the size of the catchment the whole catchment processes are described as a summation of parallel models in the gridded mode. If the grid resolution is large then the catchment processes are represented by a fraction of the response from one cell. The biggest impact of

this will be on parameters related to hill slope routing where the geometry of the catchment and potentially the drainage network are likely to have an impact on the routing of effective rainfall. This limitation is offset by the use of a monthly or longer time step in most applications of this class of model. At these resolutions the hill slope routing of runoff is not as important.

Pilling and Jones (1999) present the application of a 10-km resolution gridded application of a seasonal macro-scale model. The model was based on an application of HSYIM with *a priori* defined parameters. The approach of treating grid cells as independent entities was adopted within the model. The authors claim that this work represents a significant enhancement over the work of Arnell and Reynard (1996) both in terms of scale (true) and physical realism, which is debatable. The prediction of annual runoff by the model was evaluated against stream flow data for 865 catchments that were classified by Gustard *et al* (1992) as suitable for the regionalisation of flow statistics. The authors evaluated the model fit by averaging the grid cell predictions of annual runoff and comparing that prediction with the gauged runoff. The results are then presented as the percentage of grid cells that have annual runoff errors of less than 10%, between 10% and 20% and the percentage of cells that have errors of greater than 20%. This is incorrect, as although the simulated catchment average value of runoff may have an error of less than 10% the individual cells may well have errors greater or less than this. A more correct representation would be to present the percentage of catchments where errors are less than 10%. Holmes and Young (2000) demonstrate that it is within smaller catchments that the larger errors occur in the regional modelling of annual runoff. This is because it is in these catchments that the biggest errors in the averaging of spatial rainfall fields are observed and where inaccuracies in catchment definition have the greatest impact on the assumption of a closed water balance.

## 1.4 NEW RESEARCH AND THESIS STRUCTURE

### 1.4.1 New research

With the exception of the study of Sefton and Boorman (1997), this study is the first reported study in which the objective has been to regionalise a model for predicting daily stream flows within ungauged catchments. Other studies have concentrated on flood estimation or the prediction of monthly and/or annual-resolution stream flow data. At these longer time steps, a regionalised model has to be capable of closing an effective water balance, but the processes controlling the routing of effective rainfall through a catchment do not have to be accurately modelled as these processes tend to operate over a shorter time scale.

The study of Sefton and Boorman is inconclusive due to the relatively small sample size and the fact that the utility of the regionalised model for predicting historical stream flows was not explicitly explored within the study. The regionalisation of the model was only one aspect of the paper. The authors recognised the poor quality of some of the relationships between model parameters and catchment characteristics and the inconclusive nature of the results obtained for the regionalisation part of the study.

The research that is reported within this thesis is innovative for the following reasons.

- It is one of only two studies to focus on the regionalisation of a daily resolution, catchment scale rainfall runoff model.
- It is the largest study of this type within the UK, with nearly 180 catchments successfully incorporated into the analysis.
- It is the first study to take a holistic approach: addressing, input data errors, model structure, model calibration and parameter identifiability and regionalisation strategies.
- The results of the study have been extensively evaluated.

This study is the first to have demonstrated that, in the UK context, a conceptual model structure for predicting daily stream flows can be defined such that conceptually justifiable relationships can be derived between model parameters and the physical characteristics of the catchment being modelled. Furthermore, the stream flow simulations derived using the

regionalised model are certainly applicable for many water resource assessment applications.

### 1.4.2 Structure of the thesis

Chapter 2 presents an evaluation of methods for estimating daily spatial rainfall grids at a resolution of 1km$^2$ within the UK. The chapter also presents the derivation of a method for predicting daily spatial grids of potential evaporation demand.

Chapter 3 presents an evaluation of a range of conceptual model structures for five catchments within the Anglian region of the UK. The catchments selected are some of the driest catchments within the UK. As will be discussed in this chapter, the correct modelling of soil moisture behaviour, and the relationship between soil moisture deficits and evapotranspiration losses are crucial to the successful application of a conceptual rainfall runoff model for water resources purposes. This evaluation guided the selection of the preferred basic model structure for the research.

The development of two appropriate deterministic conceptual model structures for the regionalisation research is presented within Chapter 4. The focus of this development was to develop model structures that had a strong physical basis whilst minimising the number of parameters that were to be calibrated.

The selection of good hydrometric quality natural catchments for use within the study is presented in Chapter 5. This identified 318 candidate catchments for use within study of which nearly 180 catchments were subsequently used within regionalisation studies. The derivation of candidate catchment characteristics and descriptors for these catchments is also reported within this chapter.

The development and application of a novel, multi-objective function calibration scheme is presented within Chapter 6. A detailed analysis of the model behaviour was made over a range of catchment types using this scheme. This is illustrated within Chapter 6 with reference to four case study catchments. This analysis led to a simplified form of the initial model structure. The simplifications were made to reduce the issue of parameter covariance within the model.

The model fits obtained via the calibration procedure are evaluated across the catchment data set in Chapter 7. The development of predictive relationships between model parameters and catchment characteristics/descriptors is also presented within this chapter.

The utility of the relationships for predicting model parameters from catchment characteristics is assessed with Chapter 8. This was assessed through comparisons between the fits obtained with the models when using regionalised parameter estimates with those obtained using the calibrated model parameters. For two of the Anglian catchments, considered in Chapter 3, a comparison between the performance of the regionalised models and the models used in the original model evaluation studies is presented. These catchments were not used in the regionalisation studies.

As discussed in Section 1.1, the current practical approach to estimating stream flows at ungauged sites is based on the transposition of natural or naturalised stream flows from suitable analogue gauged catchments. The fit obtained using commonly used transposition methods is assessed in Chapter 9 for the catchments used in the regionalisation studies. A comparison is then made between the fit obtained with the best of these methods and the fit obtained through the use of one of the regionalised model forms.

The results from the studies forming this thesis are discussed in Chapter 10. This chapter presents the conclusions from the research and makes recommendations for further work.

# 2 Estimation of climatic data

The optimisation of parameters of any model will tend to compensate for measurement error within both the input data and the calibration flow data. It is therefore essential that the methods used for estimating climatic input data are both accurate and consistent in approach across the study area, in this case the United Kingdom. One objective of the research has therefore been to define and evaluate techniques for estimating time series of catchment rainfall and potential evaporation for any catchment. It has not been necessary to consider temperature explicitly, as the regionalisation of snow storage and melt processes is not part of this study. This chapter reviews the literature on the consequences of measurement errors in climatic data for hydrological modelling and describes the selection of methods for estimating areal rainfall and the development of a method for estimating potential evaporation on a daily basis and at a 1km grid resolution.

## 2.1 CONSEQUENCES OF MEASUREMENT ERRORS IN CLIMATIC DATA

Errors in the spatial fields of rainfall used to derive catchment average values may result from measurement and/or interpolation error. The standard rain gauge within the UK is the Meteorological office MKII rain gauge (Meteorological Office, 1981) set into the ground with the rim located at 300mm above the ground surface. However, increased turbulence around the gauge can lead to the gauge catch being significantly reduced. This has led to gauges being increasing set at ground level with a variety of devices to minimise in-splash. Based on measurements at Wallingford and the work of other researchers Rodda (1967) hypothesised that these error may lead to significant systematic error of about 5% in the estimation of average annual rainfall and that the degree of this systematic error might vary within the year. This is illustrated in Figure 2.1 for Rodda's Wallingford site. This data excluded months of snow, obviously precipitation that falls as snow may not be fully caught by a standard rain gauge. The issue of snow is really only a major issue in the mountainous areas of the UK and the Highlands of Scotland in particular. The analysis presented in this chapter has been restricted to assessing the potential errors of spatially interpolating point rainfall data rather the consequences of inherent systematic errors in the point measurements.

The potential for evaporation demand at a point is a derived measure based upon measurements of humidity, the energy balance and wind speed, amongst other variables. When spatially interpolating point measurements of evaporation consideration needs to be given as to whether it is better to interpolate the meteorological variables or the derived evaporation measures.

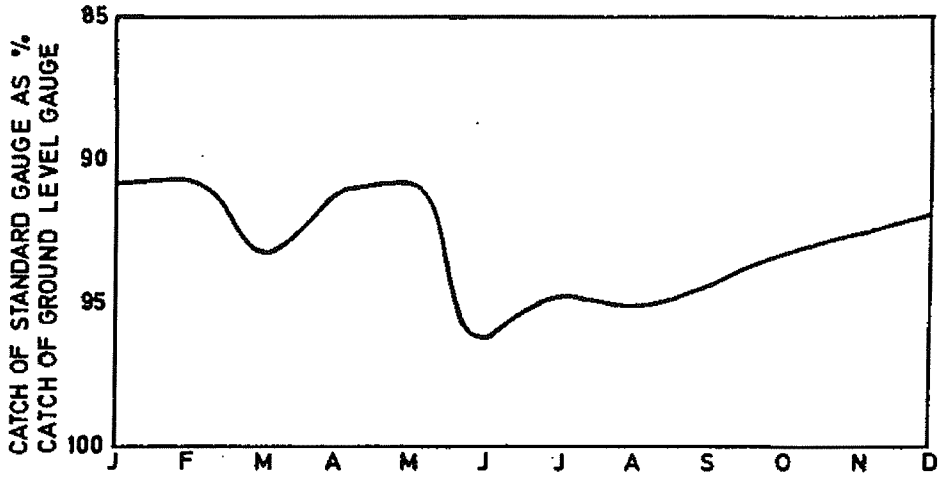



**Figure 2.1** **Mean monthly differences between ground-level and standard gauge, July 1961 to August 1966 (excluding months of snow)** ***(Source: Rodda, 1967).***

The importance of errors in rainfall fields derived from point measurements is a function of the spatial and temporal extent of the required rainfall surfaces. The type of precipitation is also important; errors are likely to be smaller for frontal precipitation rather than thunderstorms and/or localised showers associated with warm sector weather. Taking the temporal consideration the errors are likely to be higher for hourly or 15 minute data where the spatial extent of individual rain events needs to be considered. The errors for daily rainfall surfaces tend to be smaller as an average of the depth of all rainfall events within a day is taken.

Faurés *et al* (1995) looked at the implications of spatial errors in rainfall data for single events for the modelling of a small, 4.4 ha semi arid catchment using a distributed rainfall runoff model. The results of the analysis demonstrated that the model output (peak rate and total runoff volume) was extremely sensitive to the characterisation of rainfall using between 1 and 4 rain gauges. Obled *et al* (1994) evaluated the sensitivity of the TOPMODEL to the spatial characterisation of hourly rainfall. This was undertaken within a 71-km2 catchment in Southeast France over a number of events. They tested two network

densities against a baseline assumption of uniform rainfall. They concluded that the spatial variability of rainfall, although important, was not sufficiently organised in space and time to overcome the effects of catchment averaging within this size of catchment.

The majority of the literature on the impacts of errors in climatic data on stream flow simulation has been driven by the need to understand the relationship between errors in rainfall data and the ability of models to simulate catchment response to individual rainfall events. However, there are a number of studies that have looked at this issue from a resource perspective. Storm, *et al* (1988) considered the sensitivity of the NAM model to the uncertainty in catchment average daily rainfall time series derived for a Danish catchment using a kriging based interpolation technique. This analysis demonstrated that the largest errors in stream flow occurred in the winter months when the evaporation demand is smallest.

Paturel *et al* (1995) considered the sensitivity of the simple GR2M model to errors in input data in 5 medium sized catchments on the Ivory Coast of Africa. They considered the impacts of systematic (both under and over estimation) errors in monthly rainfall and evaporation demand on simulated monthly stream flow. The systematic errors in rainfall led to equivalent systematic errors in simulated stream flow with a linear relationship between the two. When expressed as a percentage error, the errors in stream flow were higher as, crudely speaking, the stream flow is the balance between rainfall and evaporation demand. The simulated stream flow was also more sensitive to percentage errors in the rainfall data than the evaporation data. They also considered the impact of random errors in the climatic data and identified that there was a non-linear relationship between the random errors and stream flow errors – a consequence of the non-linear relationship between rainfall, evaporation and resultant stream flow. Nandakumar and Mein (1997) evaluated the sensitivity of the Monash HYDROLOG model to errors in climatic data and parameter uncertainty at five experimental sites in Victoria, Australia. They again considered systematic errors in rainfall and both random and systematic errors in evaporation. The analysis identified that the simulated annual runoff was most sensitive to systematic errors (as would be expected) and that the more permeable catchments were less sensitive to the errors in the estimation of evaporation demand. The percentage error in simulated stream flow was much more sensitive to percentage errors in rainfall than evaporation.

These studies indicate that simulated stream flow is more sensitive to errors in precipitation and that the ability of a model to close a water balance (zero systematic error in simulated stream flow) is much more sensitive to systematic rather than random errors. From water balance considerations these results are intuitively correct.

## 2.2 METHODS FOR DETERMINING AREAL RAINFALL

Areal precipitation methods seek to represent the spatial distribution of precipitation over a catchment. If r(x,y,t) is taken to be the depth of precipitation at the point (x,y) within time interval (t) areal rainfall can be derived from the integral

$$\overline{R_t} = \frac{\int\int_{yx} r(x,y,t)dxdy}{\int\int_{yx} dxdy} \quad . \qquad (2.1)$$

In practice the function r(x,y,t) is not known and is thus estimated from the precipitation values measured at rain gauges. These can be regarded as point measurements across the rainfall surface r(x,y) at time, t. Most methods express the areal rainfall integral as a weighted average of the values measured at the individual gauges, expressed as

$$\overline{R}_t = \frac{1}{n}\sum_{i=1}^{n} w_{i,t} r_{i,t} \quad , \qquad (2.2)$$

where:

n = the number of gauges;

$W_i$ = the weight applied to rain gauge, $r_i$.

The weight given to a rain gauge is commonly the fraction of the catchment r(x,y,t) surface whose rainfall is represented by the rain gauge (r). The most common methods for estimating areal rainfall are domain based methods such as Thiessen polygons (Thiessen, 1911). In these methods the area of interest is subdivided into polygonal areas with a rain gauge within each. The polygons are constructed from perpendicular bisectors of nearest neighbour arcs between rain gauges. The weight for each gauge is the area of the corresponding polygon, which in turn is that part of the catchment that is closest to the gauge. This is demonstrated diagrammatically in Figure 2.2. The main disadvantages of domain based methods for areal rainfall estimation are that the gauge may not be representative of the domain and that there are discontinuities in the estimated r(x,y,t) surface at the domain boundaries. The British Standards Institute standard "Guide to the

acquisition and management of meteorological precipitation data" (British Standards Institute, 1996) recommends the use of either the triangular planes method (Jones, 1983) or Voronoi interpolation (Sibson, 1982) for generating areal rainfall estimates. These are both weighted mean methods amenable to implementation using a computer on a grid basis. The principle advantage of these methods is that they produce smooth rainfall surfaces without the boundary discontinuities that occur between adjacent polygons in the Thiessen polygon method. This is of particular importance in small catchments where domain polygons may be of a similar spatial resolution to the catchment. Within this study both the triangular Planes and Voronoi methods were adapted, implemented and evaluated for estimating daily rainfall time series for any cell within a $1km^2$ grid across the UK.

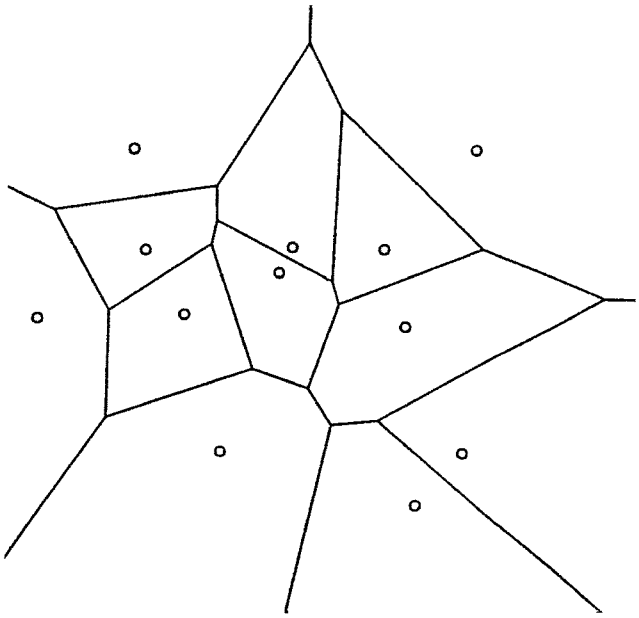


**Figure 2.2 A diagrammatic representation of Thiessen Polygons.**

The study has not considered statistical methods such as kriging (Matheron, 1963) as these are not British standard methods. These methods use rain gauge weights derived on the basis of ensuring the estimates of rainfall for each time interval at a point are both unbiased and that errors are minimum variance errors. The method assumes that these estimates are uncorrelated through time. The spatial dependency of the rainfall field is described by an auto covariance function (derived for measured points) that varies through space and through time.

Triangular planes method

The triangular planes method evaluated within this study was a simplified version of the original Jones' method which in addition to the interpolation of the point data also considered the most appropriate grid scale for application of the scheme. Taking advantage of increased computing power a standard grid of resolution $1km^2$ was used for this study.

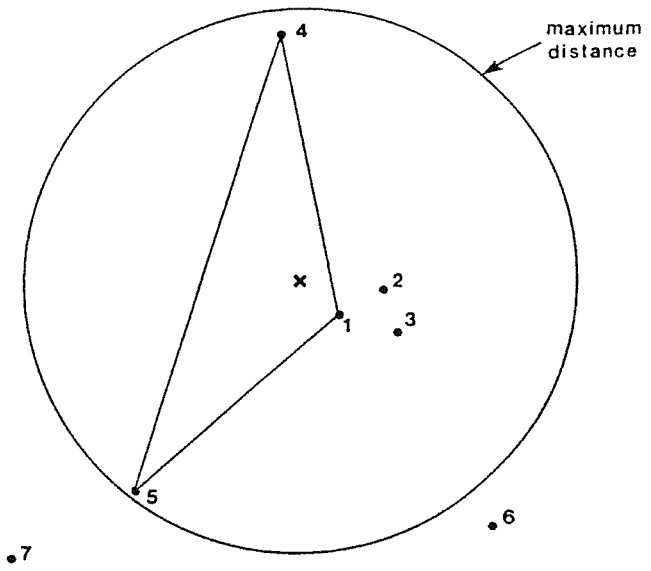



**Figure 2.3 A diagrammatic representation of the Triangular planes method.**

The starting point for the rainfall estimation procedure is a catchment boundary. The boundary is used to delineate the $1km^2$ cells from a regular grid for which daily rainfall estimates have to be derived. For each day the three closest rain gauges capable of forming an enclosing triangle are identified for each cell centroid. This is demonstrated diagrammatically in Figure 2.3. A rainfall value, rc, for a cell, i, at time, t, is then derived using a weighted average based upon the inverse distance of each gauge from the estimation point according to

$$rc_{i,t} = \sum_{j=1}^{3} w_{i,j,t} r_j \quad . \qquad \textbf{(2.3)}$$

The weight for each rain gauge $r_j$ is a Euclidean based distance measure given by

$$w_{i,j,t} = \frac{1/d_{i,j}^2}{\sum_{j=1}^{3} 1/d_{i,j}^2} \quad , \qquad (2.4)$$

where:

$w_{i,j,t}$ = Weight for rain gauge, j, in time interval, t, when used in the estimation of rainfall for target cell, i;

$d_{i,j}$ = the geographic distance between rain gauge, j, and target cell, i.

If a triangle cannot be formed, using gauges within a circle of radius 60-km from the cell centriod, the value at the nearest rain gauge is used. The areal average for the catchment is the average value for all cells subtended by the catchment boundary. This method can be further refined by normalising the measured rainfall for each gauge, $r_j$, by the Average Annual Rainfall (AAR) for the gauge and multiplying the estimated value for the target cell $rc_i$ by the average annual rainfall of the cell. This obviously requires *a priori* knowledge of the average annual rainfall for the target cell. This step minimises the impact of differences in total rainfall depth between source gauges and the target cell.

The Meteorological Office have derived a digital 1 km resolution grid of average annual rainfall for the United Kingdom at a mapped scale of 1:625,000 for the 1961-1990 standard period (Spackman, 1996). The procedure used to generate this map was based upon the derivation of node values of average annual rainfall values for a 10-km grid using monthly data from approximately 13,100 rain gauges. These values were then contoured and gridded at a 1km resolution using a bi-cubic spline interpolation procedure. Estimates of annual rainfall derived from this map (henceforth referred to as Standard Period Average Annual Rainfall, SAAR (61-90) estimates) were used to describe the AAR for each $1km^2$ cell.

Voronoi interpolation

Voronoi interpolation is essentially an extension of the Thiessen polygon approach. In Voronoi interpolation Thiessen polygons are constructed for the gauge network. Within the computer implementation, based upon a 1 km resolution grid, polygons are constructed by assigning grid cells to gauges on a closest distance basis so each gauge $r_j$ will have a set

of cells, $C_j$ associated with it. The next step is to sequentially estimate an interpolated rainfall value for each cell within the catchment boundary and for each day. For each cell, rc, the centriod of the cell is introduced as a new "rain gauge" into the network, and the Thiessen polygons re-calculated taking into account the new "rain gauge". These process results in the cell centroid, lying at the centre of a polygon which overlaps the original polygons developed for the true rain gauge network. The cell polygon consists of a set of cells, $C_{rc}$ The rainfall value for the cell is then obtained by taking a weighted average of the values at rain gauges whose original polygons are intersected by the polygon associated with the target cell. The weight assigned to a rain gauge is the fraction of the cell polygon that is overlapped by the original polygon for the gauge. The rainfall value for the cell is therefore obtained from the n gauges whose original polygons intersect the cell polygon using

$$rc_{i,t} = \frac{1}{C_{rc}} \sum_{j=1}^{n} (C_{rc} \cup C_j) r_j \quad . \qquad \textbf{(2.5)}$$

Whereas the surface produced by Thiessen polygons is a series of single value domains with discontinuities between them, the Voronoi method produces a gradually varying surface. The method can also be enhanced by use of average annual rainfall using the same approach as described for the triangular planes method.

### 2.2.1 The evaluation of rainfall estimation methods

The approach

The performance of the Voronoi interpolation and triangular planes methods was evaluated within four test areas (A, B, C & D) each covering an area of 50 $km^2$. The locations of these areas are presented in Figure 2.4. The objective of the testing was to evaluate the performance of the methods within both high and low rainfall areas of the United Kingdom. For each class two areas were selected, one with a dense rain gauge network and one with a sparse network. The characteristics of the test areas are summarised in Table 2.1. This table presents for each area the total number of rain gauges operating within the 1961-90 period, record length statistics and the Meteorological Office estimate of 1961-90 rainfall for each area.

Area A is located in East Anglia around the town of Newmarket. It was selected as a low rainfall area with a dense rain gauge network, as demonstrated by the summary statistics from Table 2.1. Area B is located over the North Pennine Moors, and represents a well-instrumented area with above average rainfall. Areas C and D are both sparsely gauged. Area C is a low rainfall area in the Grampian region of Scotland to the east of the Grampian Mountains and Area D is a high rainfall area within the Highlands region. While the total number of rain gauge years over the period 61-90 varies markedly between areas the mean record length per gauge does not.

**Table 2.1 Summary characteristics of test areas**

| | | Area | | | |
|---|---|---|---|---|---|
| | | **A** | **B** | **C** | **D** |
| **Total number of gauges** | **1961-90** | 67 | 49 | 17 | 24 |
| **Record length (yrs)** | **Max.** | 30 | 29 | 26 | 27 |
| | **Min.** | 5 | 5 | 5 | 5 |
| | **Mean** | 15 | 14 | 14 | 13 |
| **Total gauge years** | | 1025 | 667 | 244 | 313 |
| **61-90 SAAR (mm/yr)** | | 586 | 1249 | 988 | 1980 |

The performance of each of the methods was assessed through the ability of a method to replicate the observed daily rainfall pattern for each gauge within an area. Within an area, a rainfall time series was generated for each gauge over the period 1961-90 by removing the rain gauge in question and using the surrounding gauges to estimate the daily rainfall time series at the gauge site. The analysis was undertaken for the two methods both with and without enhancement through the use of SAAR data. Goodness of fit measures were used to evaluate the differences between observed and predicted time series for a gauge. These were the $R^2$ statistic and the BIAS in the simulated annual rainfall. The $R^2$ is that proportion of the variation in the observed rainfall time series explained by the variation in the simulated time series. The BIAS is given by

$$\text{BIAS} = \left(1 - \frac{\sum rs_i}{\sum ro_i}\right) \bullet 100 \,, \qquad \textbf{(2.6)}$$

where:

$rs_i$ = the simulated rainfall on day, i, within the year ($mmd^{-1}$);

$ro_i$ = the observed rainfall on day, i, within the year ($mmd^{-1}$).

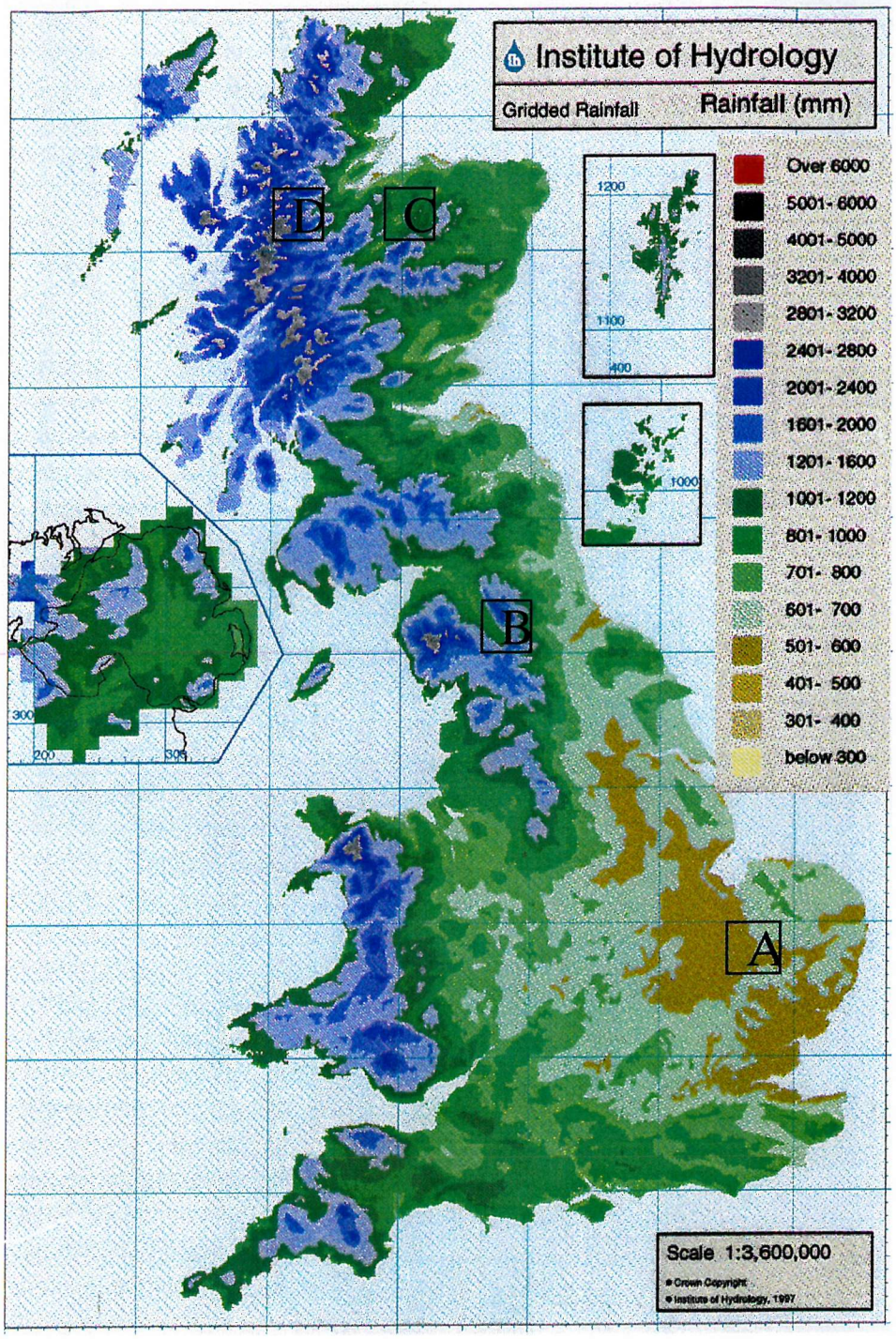



**Figure 2.4 The location of the test areas within the UK.**

For both of these measures only those days when rainfall was either observed or simulated were used. Using these analyses a time series of annual $R^2$ and BIAS statistics were obtained for each gauge within a test area. The $R^2$ statistics are a measure of the correlation between the observed and predicted rainfall time series. The BIAS statistics are a measure of whether or not the errors in daily rainfall predictions are random or whether there is a tendency for the method to over or under predict daily rainfall within a year (a systematic

error). Figure 2.5 presents an example of the BIAS time series for an example gauge in area A for both methods with and without standardisation by AAR. As can be seen the average BIAS over the 61-90 period is low however the inter year variability can be up to ± 8%.

Ideally it would have been more rigorous to select a subset of gauges from each area for evaluation purposes. Due to the low sample sizes in areas C and D this was not a practical option and thus all gauges were used for both evaluation and prediction. The consequence of this is that there will inevitably be a degree of covariance between errors at adjacent rain gauges. This is best explained by considering nearest neighbour gauges A and B. It is quite likely if gauge A is used to predict rainfall at the site of gauge B that gauge B will be used to predict rainfall at the site of gauge A.

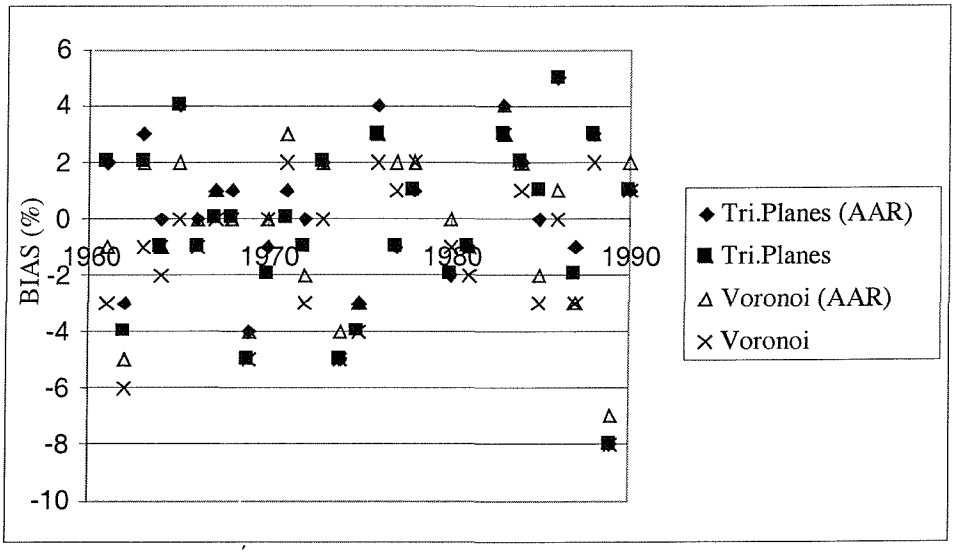



**Figure 2.5 An example of the variation in annual bias for a gauge in Area A.**

This will not always be the case, for example, at gauges near the edges of the areas where one or more of gauges used to estimate the rainfall at a gauge site may be drawn in from outside of the area. Within the Triangular planes methodology it is not axiomatic that if A is one of three nearest neighbours used to triangulate the site of gauge B that gauge B will be one of the equivalent nearest neighbours for site A. The consequence of potential error covariance between sites will be that the variance of the spatial distribution of errors at points within the areas considered will be under-estimated. It is reasonable to expect that a method that will result in poor point predictions will result in poor estimates of areal rainfall.

In practice the variation in performance between methods was found to be small. The general variation in performance between areas is therefore presented with respect to the performance of the Triangular Planes Method using the standardising by AAR (henceforth call the TPA method). The performance of the other methods is presented in the context of differences between these methods and the TPA method.

Evaluation of the differences in the performance of the TPA method between areas

To summarise the performance of methods for all gauges within an area the time series of annual statistics for each gauge were summarised as the average value for each statistic over the (61-90)-sample period and the standard deviation of the variation of the annual statistic for the sample period. Summary statistics about the variation in the value of the mean annual bias for gauges in each area and the "at gauge" variation in the annual bias statistic (as represented by the Coefficient of Variation (CV) of the annual bias statistic for each gauge) are presented in Table 2.2. The CV of annual bias at a gauge was calculated using

$$CV = \frac{\sigma_{bias}}{\overline{bias} + 100}, \qquad (2.7)$$

where:

$\overline{bias}$ = mean annual bias over the 61-90 period expressed as a percentage;

$\sigma_{bias}$ = standard deviation of annual bias over the 61-90 period.

**Table 2.2 Summary bias statistics for the TPA method**

| | | Area | | | |
|---|---|---|---|---|---|
| | | **A** | **B** | **C** | **D** |
| **Mean annual bias statistics for gauges** | **Max** | 10.1 | 10.5 | 4.9 | 18.5 |
| | **Min** | -8.6 | -18.7 | -5.4 | -9.6 |
| | **Mean** | 0.0 | 0.0 | 0.0 | 0.0 |
| **Annual bias CV statistics for gauges** | **Max** | 9.10 | 9.61 | 5.67 | 8.71 |
| | **Min** | 1.21 | 2.01 | 2.37 | 1.83 |
| | **Mean** | 3.74 | 4.83 | 4.62 | 4.79 |
| **Percentage of gauges** | **$H_0$: $\mu \neq 0$ (95% c.i.)** | 49 | 59 | 35 | 46 |

The summary statistics, presented in Table 2.2, do not identify any large differences between the performance of the TPA method in each of the areas. This is somewhat surprising given the low rain gauge density in areas C and D, and is perhaps a measure of the effectiveness of using the standardisation by AAR within the estimation procedure. If anything, Area C shows that the absolute variation in mean annual bias and annual bias CV, as represented by the maximum and minimum values, is less than for the other areas. However it should be remembered that area C has fewer rain gauges that the other areas and thus these differences are likely to be associated with differing sample sizes between areas.

A good simulation of the daily rainfall at a gauged site would have a minimal annual variance and a mean annual bias of zero. For the example gauge (Figure 2.5) it can be seen that the variation in annual bias is fairly random over the 61-90 period. This gauge was fairly representative of all gauges considered. The CV of the sample for each gauge can be taken as a first approximation to the standardised variance of the underlying population. From Table 2.2 it can be seen that the mean CV for rain gauges in each area was less than 5% and the maximum was less than 10%.

The departure from zero for the mean annual bias (indicating a systematic error) for a gauge may be a result of the sample size considered. A two tailed t test was therefore applied to the data for each gauge to test whether the mean annual BIAS for each gauge was significantly different from a ideal population mean ($\mu$) of zero, given the sample size and the variation of annual BIAS within the sample. The null hypothesis for this test was $H_0$: $\mu=0$ and the alternative hypothesis was $Ha\neq0$. The percentage of gauges in each area for which the mean annual BIAS was significantly different from zero is summarised in the last row of Table 2.2. With the exception of Area B, this percentage was less than 50%, although the variation between areas may be attributable to the different number of gauges in each area. The same analysis is presented in Table 2.3 for the annual $R^2$ statistic with the exception of the t test, which is not applicable in this case.

The summary statistics presented for $R^2$ again do not identify large differences between the performance of the TPA method in each of the areas. There is some evidence of a trend towards lower mean $R^2$ values in areas C and D, this indicates that gauge density does have

an impact upon the correct simulation of the daily variations in rainfall for this method. Considering the CV statistics for $R^2$ the range of CV is similar for all areas except for area C where the maximum value of 5% is much lower than that for the other areas. The mean values for the CV of $R^2$ are similar to those for the BIAS statistics in each area.

**Table 2.3 Summary $R^2$ statistics for the TPA method**

| | | Area | | | |
|---|---|---|---|---|---|
| | | **A** | **B** | **C** | **D** |
| **Mean annual $R^2$ statistics for gauges.** | **Max** | 0.99 | 0.96 | 0.91 | 0.95 |
| | **Min** | 0.78 | 0.79 | 0.81 | 0.79 |
| | **Mean** | 0.91 | 0.90 | 0.86 | 0.88 |
| **Annual $R^2$ CV statistics for gauges (%).** | **Max** | 7.88 | 9.81 | 9.90 | 9.99 |
| | **Min** | 1.06 | 1.17 | 2.19 | 0.95 |
| | **Mean** | 4.3 | 4.31 | 5.10 | 4.77 |

In the study areas considered it appears that the performance of the TPA method is not very sensitive to either the density of rain gauges considered or the magnitude of average annual rainfall, and the associated increased probability of greater spatial heterogeneity in rainfall patterns. The error in determining annual rainfall totals using the TPA method is primarily random in nature with approximately 50% of the gauges in an area demonstrating a small, but significant systematic error. The mean CV appears to be in the order of less than 5%. The correlation between simulated and observed rainfall is high in areas A, B and D with some evidence to indicate that the denser networks do facilitate a more accurate simulation of the daily variation in rainfall depths. Again the mean CV for $R^2$ appears to be in the order of 5%.

If the relationship between long term annual rainfall and annual runoff is strongly linear (as will be the case when evaporation rates are rarely limited by soil moisture deficits) the random errors in annual BIAS will not propagate into major errors in the prediction of long term runoff. However it should noted that BIAS errors in years where significant soil moisture deficits can potentially build up might lead to potentially large water balance errors in simulated runoff over the full period of record. Evaporation rates will be rarely limited by soil moisture deficits in the wetter catchments in the north and west of the United Kingdom but will occur in most years in drier parts of the country.

Comparison of the performance of the TPA method with that of the other methods

To determine whether differences between methods as determined by mean annual BIAS were significantly different from that of the TPA method two tailed t-tests were applied to each gauge using the estimates of mean annual BIAS and the standard deviation of annual BIAS for each method. The null hypothesis tested was Ho: $\mu_1=\mu_2$, where $\mu_1$ is the mean annual BIAS for the TPA method and $\mu_2$ the mean annual BIAS for the method being tested. Where the t-test demonstrated that differences were significant the differences were grouped according to whether the mean annual BIAS for the method being tested was greater or smaller than that obtained using the TPA method.

The results from these tests are presented in Table 2.4. The results demonstrate that, where significant differences exist between the methods, the differences are for those methods which do not normalise by AAR, and are invariably a consequence of larger BIAS values than those observed for the TPA method. The numbers of gauges where values are significantly different represent a relatively small percentage of the total number of gauges within the area except for the Voronoi method in area C. This result is probably a consequence of the small gauge sample size for C. The mean BIAS values are not significantly different for any of the methods for gauges in area A and only for the Voronoi method without normalisation by AAR in Area C.

In the wetter areas, B and D, instances of gauges where predictions are significantly different if normalisation by AAR is not used are observed. In all cases these result from a poorer simulation of annual rainfall (and hence higher mean BIAS) that that obtained using the TPA method. This result implies that the value of standardising by AAR increases as the mean AAR within an area increases and the associated spatial heterogeneity increases. The association between the magnitude of annual rainfall and spatial heterogeneity is illustrated within Figure 2.4.

The variation in CV of annual BIAS relative to that for the TPA method is summarised in Table 2.5 in which the subscripts 1 and 2 respectively refer to the TPA method and the method being evaluated. Here t-tests were not applied.

This table generally demonstrates that more gauges within an area have a higher CV for the Triangular planes based methods than for the Voronoi based methods, particularly for Voronoi normalised by AAR. It important to note that the improvement is less than 2% and that where the TPA method performs better the advantage the TPA method has is generally larger. However, all of these differences must be considered as being marginal.

**Table 2.4 Variations in mean annual BIAS between methods**

| | | % of gauges | | | mean % differences | | |
|---|---|---|---|---|---|---|---|
| | | Tri.Planes | Voronoi | Voronoi | Tri.Planes | Voronoi | Voronoi |
| | Area | no AAR | AAR | No AAR | no AAR | AAR | no AAR |
| **Area A** | | | | | | | |
| $\mu_1 \neq \mu_2$ | $\mu_2<\mu_1$ | 0 | 0 | 0 | | | |
| | $\mu2>\mu1$ | 0 | 0 | 0 | | | |
| $\mu_1=\mu_2$ | | 100 | 100 | 100 | | | |
| **Area B** | | | | | | | |
| $\mu_1 \neq \mu_2$ | $\mu_2<\mu_1$ | 0 | 0 | 0 | | | |
| | $\mu2>\mu1$ | 6 | 0 | 8 | -25.13 | | -19.95 |
| $\mu_1=\mu_2$ | | 94 | 100 | 92 | | | |
| **Area C** | | | | | | | |
| $\mu_1 \neq \mu_2$ | $\mu_2<\mu_1$ | 0 | 0 | 0 | | | |
| | $\mu2>\mu1$ | 0 | 0 | 18 | | | -15.53 |
| $\mu_1=\mu_2$ | | 100 | 100 | 82 | | | |
| **Area D** | | | | | | | |
| $\mu_1 \neq \mu_2$ | $\mu_2<\mu_1$ | 0 | 0 | 0 | | | |
| | $\mu2>\mu1$ | 4 | 0 | 4 | -10.40 | | -10.60 |
| $\mu_1=\mu_2$ | | 96 | 100 | 96 | | | |

The same analysis undertaken for annual BIAS was also under taken for the $R^2$ statistics. The analysis for $R^2$ demonstrated that in all cases the performance of the methods in all cases was not significantly different. Furthermore differences between the CV values for $R^2$ were in all cases less than 1%.

The analysis has demonstrated that there is little to choose between the Triangular planes method and the Voronoi interpolation methods when the rainfall estimates from the source gauges are normalised by the average annual rainfall for the gauge prior being used in the either of the weighted averaging schemes. The standardisation process appears to offer best improvements in wetter areas, in terms of BIAS. The temporal variation in simulated daily rainfall, as represented by the mean annual $R^2$ value, tends to be insensitive to the method

used although the methods do appear to simulate daily variations more effectively in drier areas and where the rain gauge network is more dense. This is intuitively correct, as the spatial heterogeneity in rainfall is lower over these dryer, lower elevation areas.

**Table 2.5 Variations in the CV of annual BIAS between methods**

| | | Percentage of gauges | | | mean % differences | | |
|---|---|---|---|---|---|---|---|
| | Area | Tri.Planes no AAR | Voronoi AAR | Voronoi No AAR | Tri.Planes no AAR | Voronoi AAR | Voronoi no AAR |
| **Area A** | $CV_2<CV_1$ | 36 | 51 | 51 | 0.18 | 0.36 | 0.39 |
| | $CV_2>CV_1$ | 63 | 46 | 49 | -0.19 | -0.26 | -0.32 |
| | $CV_2=CV_1$ | 1 | 3 | 0 | | | |
| **Area B** | $CV_2<CV_1$ | 33 | 59 | 45 | 0.98 | 0.83 | 1.19 |
| | $CV_2>CV_1$ | 67 | 41 | 55 | -3.23 | -0.42 | -2.24 |
| | $CV_2=CV_1$ | 0 | 0 | 0 | | | |
| **Area C** | $CV_2<CV_1$ | 41 | 76 | 53 | 0.21 | 1.11 | 1.20 |
| | $CV_2>CV_1$ | 59 | 24 | 47 | -1.36 | -1.22 | -1.17 |
| | $CV_2=CV_1$ | 0 | 0 | 0 | | | |
| **Area D** | $CV_2<CV_1$ | 29 | 67 | 38 | 0.92 | 0.69 | 1.09 |
| | $CV_2>CV_1$ | 71 | 33 | 63 | -2.51 | -1.13 | -1.95 |
| | $CV_2=CV_1$ | 0 | 0 | 0 | | | |

### 2.2.2 The influence of standardising by AAR on catchment estimates of average annual rainfall

The Triangular planes method is more computationally efficient than the Voronoi based methods and thus Triangular planes method enhanced through the standardisation by AAR was used to generate daily rainfall series for all catchments used within this study (see Chapter 5). The 61-90 average annual rainfall estimates generated for each catchment using the triangular planes method, with and without additional enhancement through the use of AAR are plotted as a function of Meteorological Office 1961-90 SAAR in Figure 2.6. The strong correspondence between estimated AAR generated using the TPA, incorporating the normalisation by AAR, and the published 61-90 data is not surprising as this estimate of AAR is used to re-scale the normalised estimated rainfall in the method. However the plot demonstrates that the 61-90 SAAR for high rainfall catchments is consistently underestimated when AAR is not included within the estimation procedure. High rainfall catchments lie in the west and north of the UK in which the high rainfall is a consequence

of orographically enhanced precipitation arising from rainfall associated with depressions. Furthermore the spatial heterogeneity in rainfall is higher within the areas with higher topographic relief. More that 60% of the annual rainfall in the UK is associated with depressions (Shaw, 1988). This systematic under estimation in high rainfall catchments demonstrates that the siting of rain gauges tends to be biased towards accessible low altitude areas. From a simplistic viewpoint these areas are either coastal or areas in which rainfall is generally lower as a consequence of rainfall shadowing.

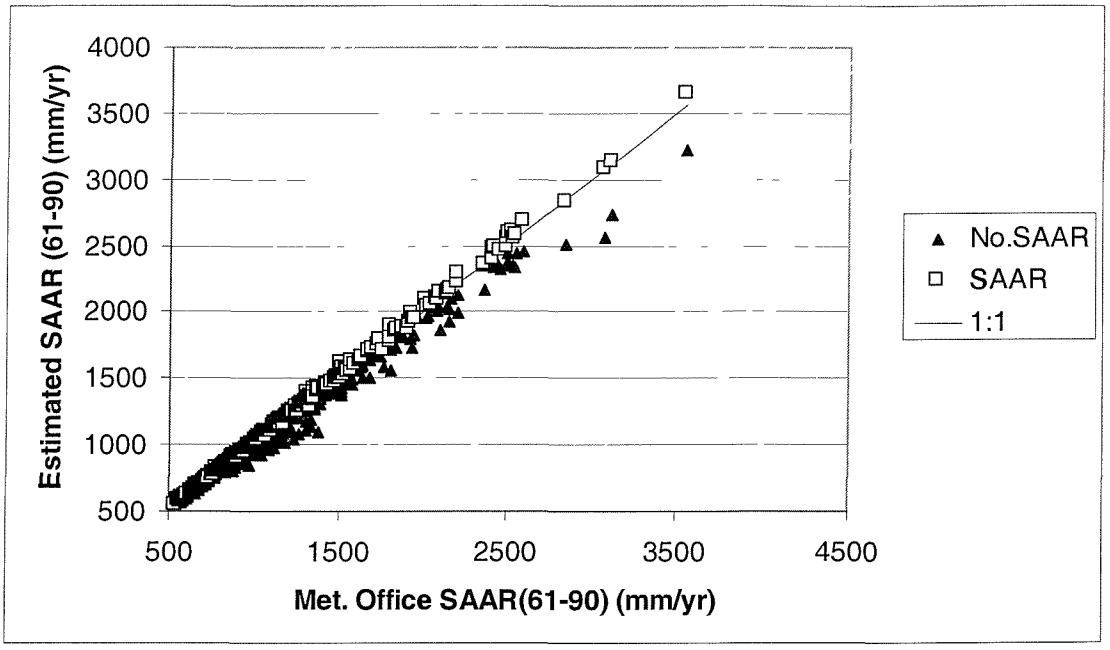



**Figure 2.6 1961-90 AAR estimates, derived using the Triangular Planes method, with and without standardisation by SAAR (61-90), and plotted as a function of SAAR.**

## 2.3 ESTIMATION OF POTENTIAL EVAPORATION TIME SERIES

### 2.3.1 National UK Potential Evaporation Estimation: MORECS

The term "potential" evaporation can be equated to an upper limit to evaporation in a given environment. Within the rainfall runoff model potential evaporation is treated as an intermediate parameter in the estimation of actual evaporation. This is obtained by reducing the potential evaporation estimate in proportion to soil moisture deficit as described in Chapter 4. In this situation, the potential evaporation estimates need to provide a correct, spatially and temporally consistent reference on which the actual evaporation scheme can subsequently be based. It is essential therefore that the physical basis for these estimates must be consistent with evaporation theory.

Calder *et al* (1983) and Anderson and Harding (1991) have suggested that using a single value of potential evaporation applied across the entire United Kingdom (but varying with time) does not produce large errors in the estimation of soil moisture deficits, with the exception of mountainous areas. This would imply that the spatial variation of meteorological variables is not of key importance. However, these studies used a range of actual evaporation estimation schemes to simulate changes in soil moisture deficit evaluated against those estimated experimentally using a neutron probe.

The Meteorological Office Rainfall and Evaporation Calculation System (MORECS) (Hough, 1996) is the only consistent, national model for estimating historical potential evaporation in the UK. Potential evaporation estimates within MORECS are based upon the Penman-Monteith equation (Monteith, 1965) and are output on a 40km resolution grid basis. The use of a physically based equation, such as Penman-Monteith is likely to give a more consistent basis for PE estimation and thus MORECS PE estimates were used as a basis for the current study. The biggest limitation of the MORECS system for PE estimation is the spatial resolution of the system. This limitation was addressed during the course of this study.

The origins of the Penman Monteith equation can be traced back to the Penman (1948) equation for evaporation from open water:

$$E_o = \frac{\Delta}{\gamma} A + \frac{E_a}{\left(\frac{\Delta}{\gamma} + 1\right)} \quad , \qquad (2.8)$$

where:

$E_o$ = evaporation from open water (mm $day^{-1}$);

$\Delta$ = slope of the curve of saturation vapour pressure with temperature (Pa $K^{-1}$);

A = available energy (mm $day^{-1}$);

$E_a$ = $f(u)(e_a-e_d)$;

u = wind speed (m $day^{-1}$);

$e_a$ = saturated vapour pressure at the ambient temperature (Pa);

$e_d$ = saturated vapour pressure at the dew point (Pa), where the dew point is the temperature at which a parcel of air must be cooled for it to become saturated;

$\gamma$ = psycrometer constant (66 Pa $K^{-1}$ at Standard Temperature and Pressure).

Penman adapted the wind function in light of further experimental evidence to give direct estimates of potential transpiration (Penman, 1963). In this equation potential evaporation is defined as "a measure of the transpiration rate from an extensive short green cover, completely shading the ground and adequately supplied with water". The equation for potential evaporation is

$$E_{tg} = \frac{\Delta}{\Delta+\gamma}[R_s(1-\alpha)] - \frac{\Delta}{\Delta+\gamma}R_{nl} + \frac{\gamma}{\Delta+\gamma}f(u)(e_a - e_d), \qquad \textbf{(2.9)}$$

where:

$R_s$ = solar radiation (mm $day^{-1}$);

$\alpha$ = surface albedo;

$R_{nl}$ = nett long wave radiation (mm $day^{-1}$).

Both the solar radiation and net long wave radiation are determined from empirical formulae, the former based on the number of hours of bright sunshine and the latter based on sunshine hours, air temperature and vapour pressure. Allen *et al* (1994) describes equations for the calculation of the Penman variables.

The Penman-Monteith equation is based on the Penman model but incorporates a revised physical representation of evaporative water loss from by including vegetation effects as defined by a "surface resistance" to represent the resistance to the diffusion of water vapour from the intercellular spaces of the leaves to the atmosphere.

The Penman-Monteith equation given by

$$E_t = \frac{\Delta(R_n - G) + \rho c_p (e_a - e_d)/r_a}{\lambda(\Delta + \gamma(1 + r_s/r_a))}, \qquad \textbf{(2.10)}$$

where:

$E_t$ = evaporation rate (kg $m^{-2}$ $day^{-1}$);

$R_n$ = nett solar radiation (kg $m^{-2}$ $day^{-1}$);

$G$ = soil heat flux (kg $m^{-2}$ $day^{-1}$);

$\rho$ = density of air (kg $m^{-2}$);

$c_p$ = specific heat of air at constant pressure (J $kg^{-1}$ $K^{-1}$);

$r_a$ = nett resistance to water vapour diffusion from the surface to the height of the measurement instrument (day $m^{-1}$);

$\lambda$ = latent heat of vaporisation of water (J $kg^{-1}$);

$r_s$ = nett resistance to water vapour diffusion from leaf and soil surfaces (day $m^{-1}$).

MORECS is a discrete grid based lumped model. Within MORECS the evaporation from (potentially) 14 land cover categories from each of three different soil types (virtual soils with a high, medium and low available water capacity) is calculated. Although MORECS uses the Penman-Monteith equation it does not calculate specific reference evaporation rates for individual land cover categories. The MORECS potential evaporation is the Penman-Monteith evaporation from grass that is freely supplied with water.

The meteorological data used to run the MORECS model is supplied by a network of 59 climate stations reporting sunshine hours and 156 stations additionally reporting rainfall, wind speed, humidity and temperature. The interpolation procedure used within MORECS is as follows. Daily averaged station data are, initially, normalised by:

- converting sunshine hours into a percentage of the mean monthly number of hours;
- reducing temperature and vapour pressure measurements to sea level values;
- standardising the wind speed using an empirical factor related to terrain roughness.

The nearest nine stations to a MORECS grid cell are then selected as long as they are within 100 km of the centre of the square. If there is a station within 0.5 km of the square centre then its measurements are used alone. The nine stations selected are reduced to a maximum of six dependent on data availability and excluding stations where there are more than two in a single octant of the grid square. If less than three stations have been identified in this way then inverse distance weighted averaging is used, otherwise plane fitting is employed. The values calculated in this way are then de-normalised. Hough *et al* (1996) state that these procedures will produce estimates of temperature, humidity and wind speed within acceptable error limits as their spatial variation is small. The sparseness of the sunshine recorders and the poor representation of daily rainfall, however, lead to less accurate estimates of these variables. The reference PE estimates generated by MORECS are not influenced by the poor rainfall characterisation within MORECS.

The Meteorological Office operates MORECS as a commercial system. Although the model runs on a daily time step the output is available at either a weekly or monthly resolution. The data used in this study was restricted to monthly data due to the prohibitive cost of the weekly data. The primary limitation of MORECS, as a source of historical PE data, is the grid resolution of 40km. The impacts of this are twofold; the scale means that the majority of the spatial variations in PE associated with variations in elevation and wind speed are averaged out and the discontinuities between adjacent cells are significant at the scale of catchments modelled within the study. The variables in the Penman Monteith equation that are sensitive to temperature and/or pressure variations related to elevation are the:

- slope of the curve of saturation vapour pressure with temperature( $\Delta$ )(Pa $K^{-1}$),
- wind speed, $u$, (m $day^{-1}$);
- saturated vapour pressure at the ambient temperature, $e_a$, (Pa);
- saturated vapour pressure at the dew point temperature, $e_d$, (Pa);
- psycrometer constant, $\gamma$, ( Pa $K^{-1}$);
- net radiation ( $R_n$) (kg $m^{-2}$ $day^{-1}$).

A temperature lapse rate of -0.6°C/100m and a saturated vapour pressure lapse rate of -0.0025kPa/100m are used within MORECS (Hough, 1996). The importance of adjusting

for elevation is further highlighted by experimental evidence from the Balquidder catchments (Wright and Harding, 1993) that indicates grass evaporation ceases at the low temperatures encountered in the Highlands of the UK. A study carried out by Beven (1979) showed that estimates of actual evaporation made using the Penman-Monteith equation are highly sensitive to all other meteorological inputs (and especially radiation measurements). However, this sensitivity was found to be far less important than the sensitivity of the equation to changes in vegetation as mediated by the resistance values. This is not a factor within this study as the PE estimates are being used.

For this study an approach was developed for disaggregating the 40-km resolution MORECS grid estimates to a 1-km grid empirically taking into account taking in to account the effect of sub-MORECS cell scale variations in elevation. This has been achieved through the development of lumped lapse rates that encapsulate the meteorological dependencies on elevation.

### 2.3.2 Enhancement of MORECS Potential Evaporation estimates taking into account spatial heterogeneity related to elevation

A generalised 1-km resolution grid of mean cell elevation has been derived at the Centre for Ecology and Hydrology - Wallingford from the Ordnance Survey 50-m grid of elevation data. This grid was used as the basis for the spatial interpolation of MORECS PE estimates to a 1-km grid accounting for spatial heterogeneity in elevation. Empirical PE lapse rates were derived and used in an interpolation scheme based upon the triangular planes method of Jones used for rainfall estimation.

Derivation of mean monthly lapse rates

Empirical monthly Lapse Rates (LR) for PE were derived by analysing the within month variations in PE between MORECS grid cells as a function of elevation. The Penman Monteith equation is sensitive to the impact of latitude variations on nett radiation estimates, which is one of the primary variables affected by changes in elevation. To identify whether there were any latitude variations in monthly PE lapse rates the variation in monthly PE as a function of elevation was investigated across several "strips" of MORECS cells across regions where there is a reasonable variation in elevation. Those

considered were cells; 33-39, 60-66, 75-80, 83-87, 111-121 and 143-153. The MORECS grid structure is presented in Figure 2.7.

For each of the six "strips" the mean (61-90) PE was estimated for January (low PE) and July (high PE) for each cell within a strip. Lapse rates were estimated from linear regressions of PE against elevation for each month and for each strip; in this context the lapse rate is the gradient of the relationship. Lapse rates for each strip considered are presented in Table 2.6 and example plots of the relationships for strips with Northings 760 km and 320 km respectively are presented in Figure 2.8. The data show that lapse rates are generally higher in the summer months, a function of the increased nett radiation and the elevation dependency of nett radiation. However, there is no firm evidence of a relationship between lapse rate and latitude dependent variations in nett radiation. This indicates that any latitude effects are being masked by uncertainties introduced by the small and different sized cell samples between strips and the varying degrees of elevation variation across strips.

**Table 2.6 Predicted January and July lapse rates for example strips of MORECS cell with constant Northings**

| Strip | Lapse Rates (mm/m) | |
|---|---|---|
| Northing | January | July |
| 760 | -0.0093 | -0.0142 |
| 600 | -0.0202 | -0.0164 |
| 520 | -0.0101 | -0.0086 |
| 480 | -0.0112 | -4.0036 |
| 320 | -0.0086 | -6.0001 |
| 200 | -0.0064 | -4.0010 |

PE data from all 190 MORECS cells were consequently used to derive the final monthly lapse rates. The frequency distribution of mean cell elevation for all MORECS cells is presented in Figure 2.9. This demonstrates the skew to lower elevations and that the highest mean cell elevation is less than 560 m, which demonstrates the averaging effect of the 40-km resolution.

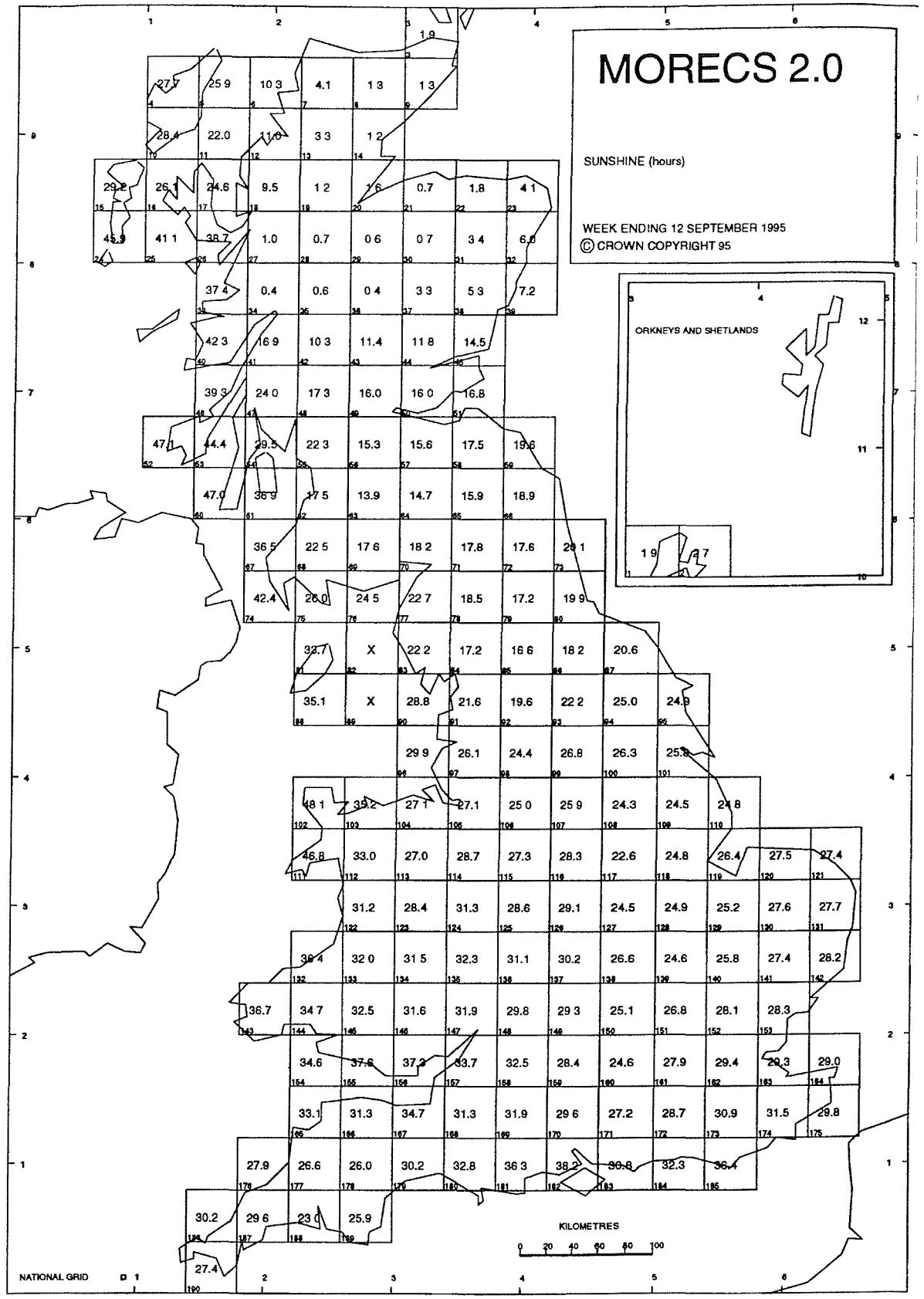


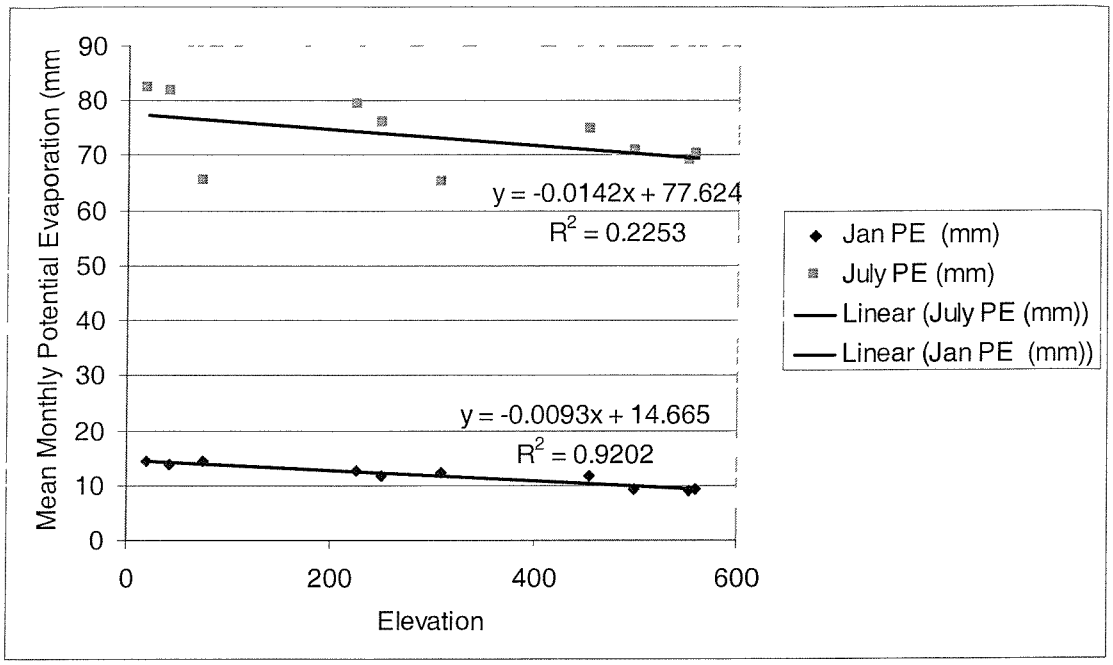

**Figure 2.7** **MORECS 2.0 grid structure annotated with sunshine hours, week ending 12th September 1995** ***(Source: Hough, 1996, © Crown Copyright 95).***

Mean monthly lapse rates were derived for each month using linear regression to relate the rate of change of mean 1961-90 monthly PE estimates to changes in elevation. As examples the relationships are plotted for January and July in Figure 2.10a and Figure 2.10b respectively. The annual PE relationship is plotted in Figure 2.10c for comparison. The estimated lapse rates are presented for all months in Table 2.7 and are presented graphically in Figure 2.11. The standard error for the individual lapse rates are also presented in Table 2.7 with the associated 95% upper and lower confidence limits, these are also presented in Figure 2.11.

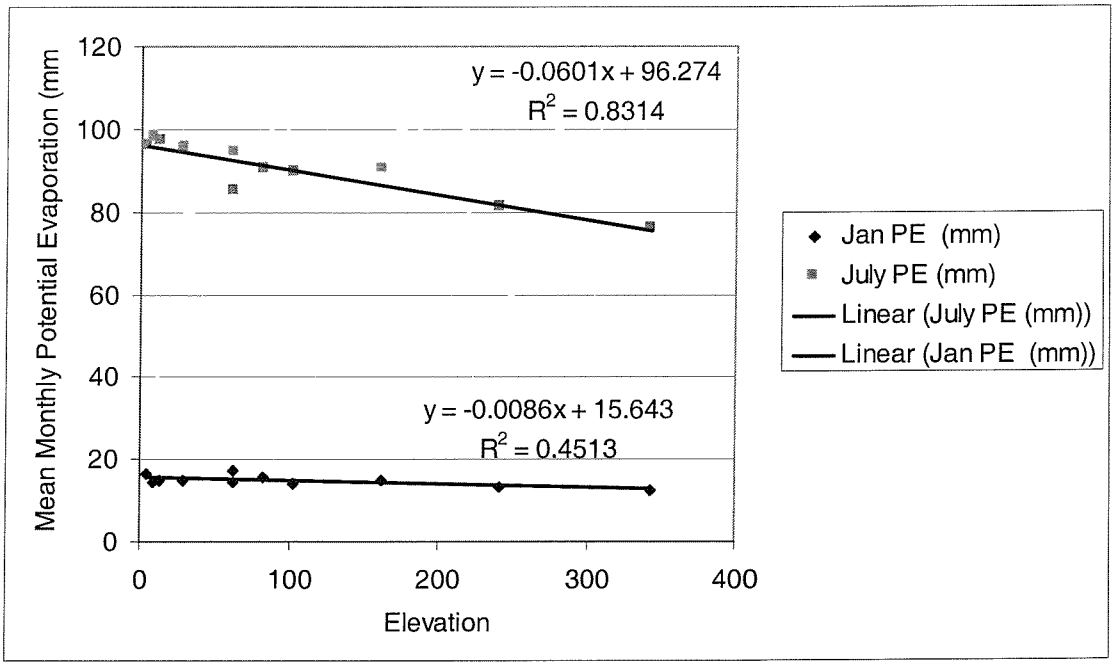
The largest confidence intervals are associated with the lapse rates during the summer months. This is related to the impact of elevation variation on nett radiation which is high in summer and which will exhibit a relatively large, climatically driven inter year variability. One limitation of the derived lapse rates is that the maximum elevation in the data set is 558m and thus when the lapse rates are applied to elevations greater than this there will be a degree of uncertainty. In practice PE is generally much lower than rainfall in these areas and thus the impact on the performance of the rainfall runoff model will be less sensitive to errors in PE in these areas than those in low rainfall areas.

**Table 2.7 Derived mean monthly lapse rates**

| | **Lapse rate (mm $m^{-1}$)** | **Standard Error (mm $m^{-1}$)** | **Lower 95% c.l (mm $m^{-1}$)** | **Upper 95% c.l (mm $m^{-1}$)** |
|---|---|---|---|---|
| **January** | -0.0143 | 0.0012 | -0.0166 | -0.0120 |
| **February** | -0.0140 | 0.0009 | -0.0158 | -0.0122 |
| **March** | -0.0180 | 0.0015 | -0.0209 | -0.0150 |
| **April** | -0.0237 | 0.0024 | -0.0284 | -0.0191 |
| **May** | -0.0344 | 0.0038 | -0.0418 | -0.0269 |
| **June** | -0.0314 | 0.0046 | -0.0404 | -0.0224 |
| **July** | -0.0388 | 0.0061 | -0.0509 | -0.0268 |
| **August** | -0.0411 | 0.0051 | -0.0511 | -0.0311 |
| **September** | -0.0316 | 0.0028 | -0.0371 | -0.0262 |
| **October** | -0.0225 | 0.0017 | -0.0258 | -0.0191 |
| **November** | -0.0177 | 0.0015 | -0.0207 | -0.0147 |
| **December** | -0.0136 | 0.0012 | -0.0161 | -0.0112 |
| **Annual** | -0.3011 | 0.0276 | -0.3557 | -0.2466 |




**(a) MORECS Cells at Northing=760km**




**(b) MORECS Cells at Northing=320km**

**Figure 2.8 Examples of the variation in PE with elevation across MORECS cells with constant Northings.**

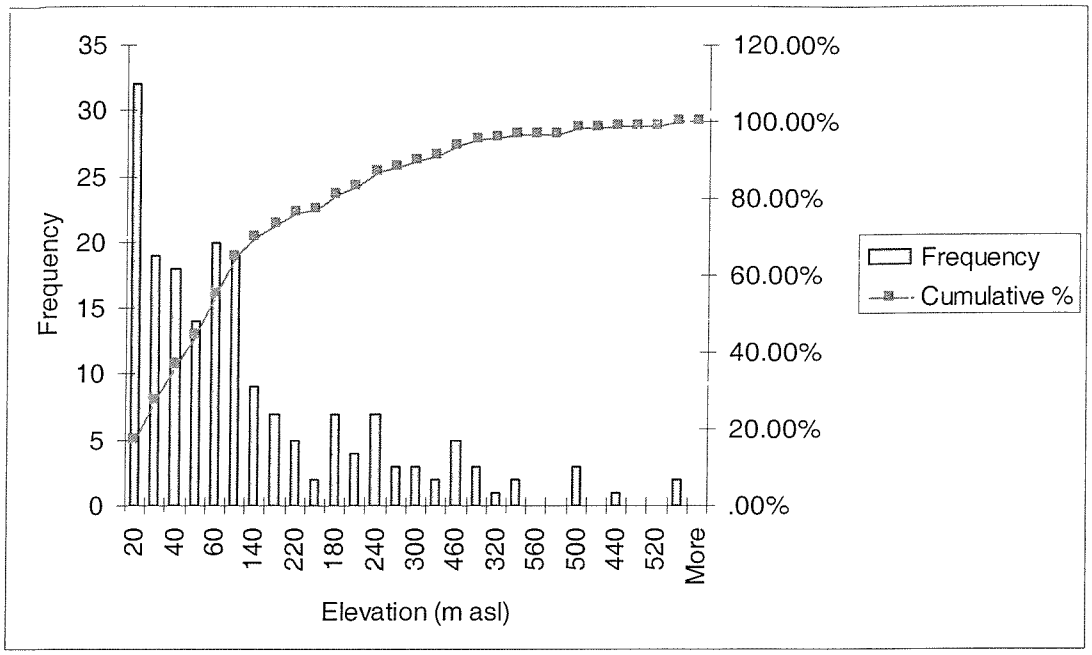
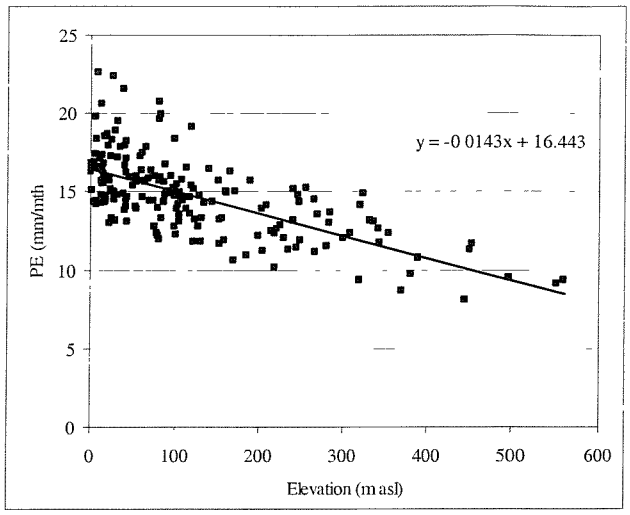


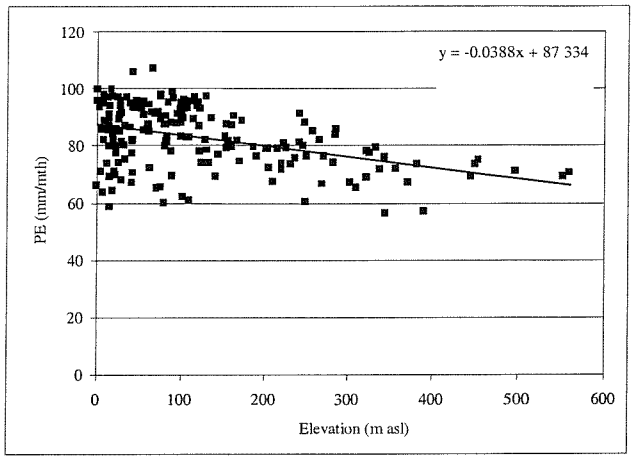

**Figure 2.9 Histogram of mean cell elevation for all MORECS cells.**

### 2.3.3 The interpolation scheme

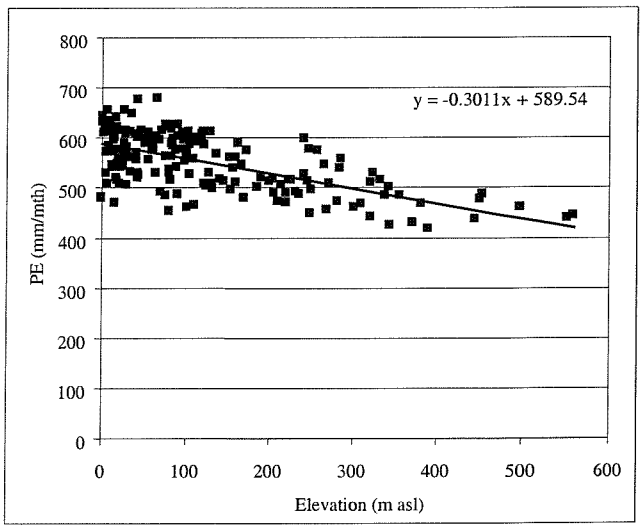
An inverse distance-weighting scheme was used to interpolate the MORECS monthly values from the 40-km resolution grid to a 1km-resolution grid. In the scheme the PE estimate and mean cell elevation for each MORECS cell are associated with the cell centriod. When interpolating to a 1 km cell, the four MORECS cells (or three at the extent of the coverage) that define a minimum area box containing the target cell are identified. The time series of monthly PE estimates for each MORECS cell are then re-scaled to the elevation of the target cell by using the appropriate monthly lapse rates in conjunction with the difference in elevation between the centriod of the target cell and that of the source MORECS cell. A weighted average of the re-scaled PE time series from the four MORECS cell is then taken. The inverse distance weight used in this average is a two dimensional Euclidean distance weight. The first component of the weight is the square of the distance between the MORECS grid cell centroids and the centroid of the target cell. The second component is the square of the difference in average elevation of the MORECS cell and the target cell. The weight is then the reciprocal of the product of these two components normalised by the sum of this reciprocal for the MORECS cells being considered.




**(a) January**




**(b) July**




**(c) Annual**

**Figure 2.10 Relationships between PE, the month and elevation for all MORECS cells over the period 1961.**

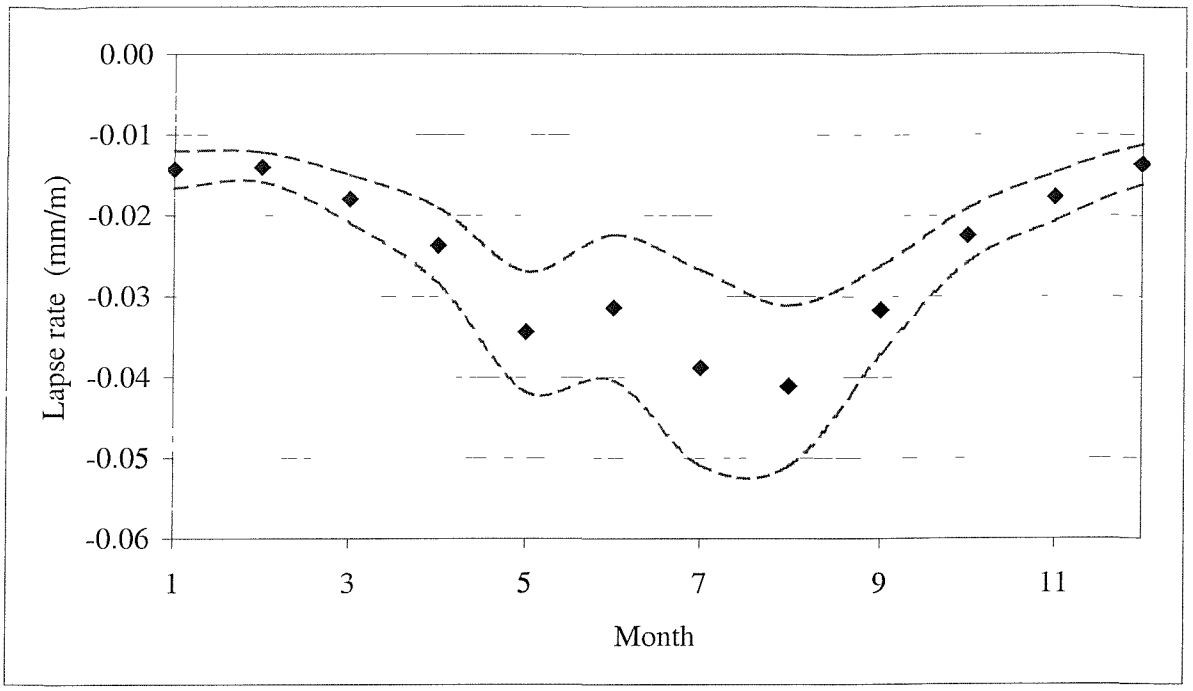



**Figure 2.11 Derived mean monthly Lapse Rates.**

The is expressed algebraically as

$$w_{i,j} = \frac{1/(d_{i,j}^2 h_{i,j}^2)}{\sum_{j=1}^{n} 1/(d_{i,j}^2 h_{i,j}^2)} \quad , \qquad \textbf{(2.11)}$$

where:

$w_{i,j}$ = 2D inverse Euclidean distance weight for MORECS cell, j, with respect to target cell, i;

$d_{i,j}$ = geographic distance between target cell, i, and MORECS cell ,j;,

$h_{i,j}$ = difference in mean elevation between the target cell, i, and MORECS cell, j.

The PE estimate for a cell, i, in month, m, is estimated from

$$PEC_{i,m} = \sum_{j=1}^{4} w_{i,j} PE'_{j,m} \quad , \qquad \textbf{(2.12)}$$

where $PEC_{i,m}$ is the PE estimate for cell, i, in month, m and PE'j,m is the height adjusted PE for MORECS cell, j.

$PE'_{j,m}$ is given by

$$PE'_{j,m} = PE_{j,m} + LR_m \cdot \left( h_i - h_j \right), \qquad \textbf{(2.13)}$$

where:

$PE_{j,m}$ = MORECS PE estimate for cell j in month, m;

$LR_m$ = lapse rate for month, m;

$h_i$ = mean elevation of the 1-km target cell;

hj = mean elevation of the 40 km MORECS cell.

The implications for catchment estimates of annual PE demand.

Figure 2.11 presents a 1-km resolution grid of 1961-90 average annual PE derived from the interpolated monthly time series of PE for each cell over the period. The percentage differences between the MORECS 61-90 estimates of annual PE and the interpolated estimates of annual PE within the catchments selected for this study are presented as a histogram in Figure 2.13a. This demonstrates that in most of the catchments the catchment PE estimates derived from the interpolated PE data are lower than those derived using the raw MORECS data. The percentage differences are plotted as a function of catchment elevation in Figure 2.13b which demonstrates a strong positive relationship between the size of the difference and the mean catchment elevation. This demonstrates the advantage gained through the use of the 1km interpolation procedure in deriving more realistic PE estimates, particularly for higher elevation catchments.

## 2.4 SUMMARY

A comparison of the two British Standards Institute methods for estimating daily rainfall surfaces was undertaken over a range of climatic conditions within the United Kingdom and for low and high density monitoring networks. The analysis has demonstrated that there is little to choose between the Triangular planes method and the Voronoi interpolation methods when the rainfall estimates from the source gauges are normalised by the average annual rainfall for the gauge prior to being used in the either of the methods. The normalising process appears to offer best improvements in wetter areas, in terms of BIAS. The temporal variation in simulated daily rainfall, as represented by the mean annual $R^2$ value, tends to be insensitive to the method used although the methods do appear to

simulate daily variations more effectively in drier areas and where the rain gauge network is more dense.

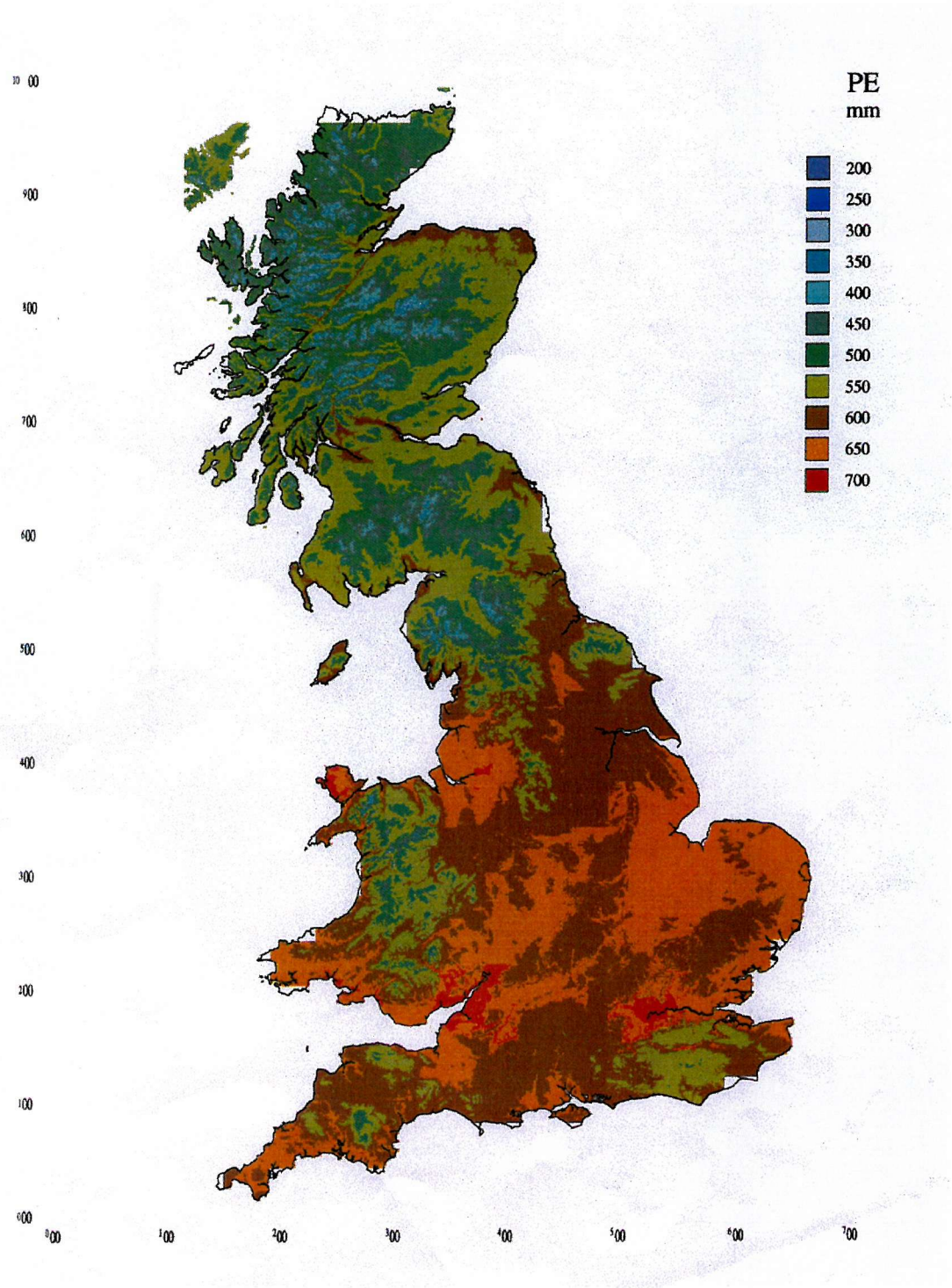


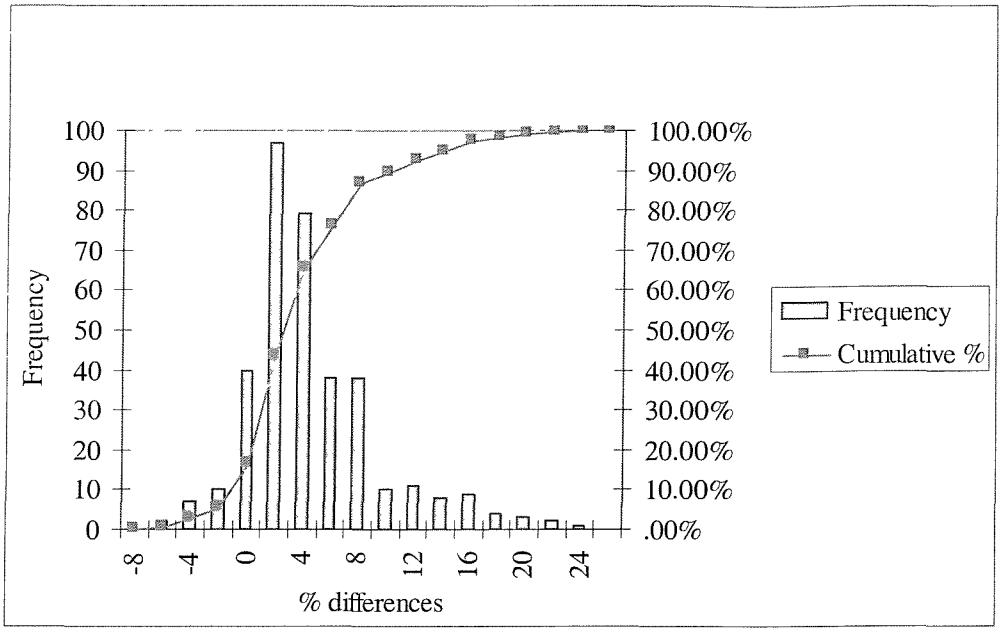

**Figure 2.12 The 1-km standard period (1961-90) average annual potential evaporation grid based on MORECS Penman Monteith estimates for short grass.**

The Triangular planes method is more computationally efficient than the Voronoi based methods and thus Triangular planes method enhanced through the normalisation by AAR (TPA) has been used for to generate annual rainfall time series for all of the catchments used in the research.

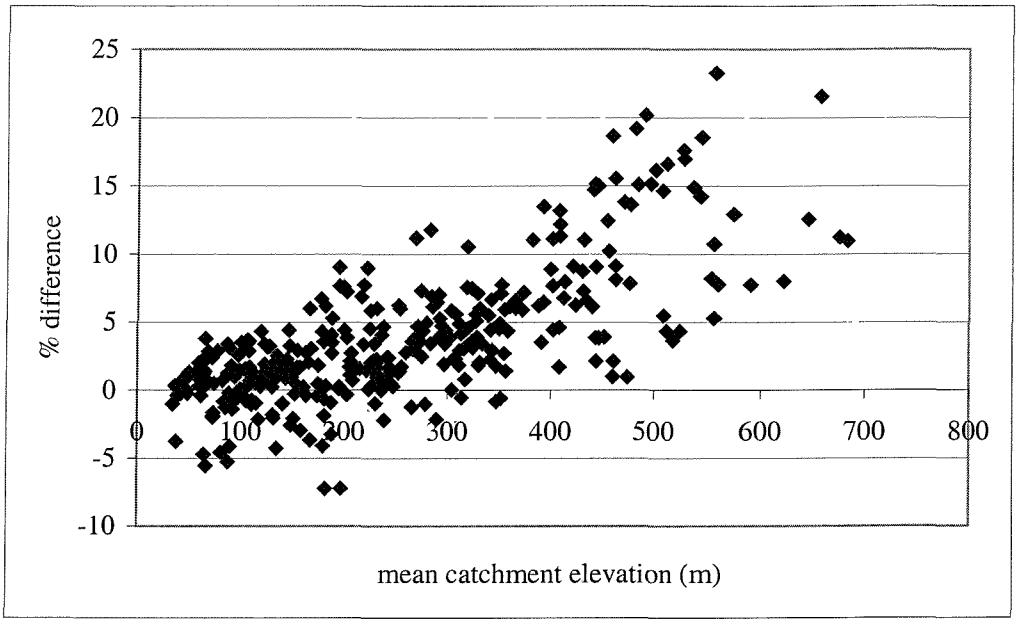
In the study areas considered it appears that the performance of the TPA method is not very sensitive to either the density of rain gauges considered or the magnitude of average annual rainfall, and the associated increased probability of greater spatial heterogeneity in rainfall patterns. The errors in determining annual rainfall using the TPA method are primarily random in nature with approximately 50% of the gauges in an area demonstrating a small, but significant systematic error. The mean CV for the variation in annual bias for individual gauges is in the order of less than 5%. The correlation between simulated and observed is high with some evidence to indicate that the denser networks do facilitate a more accurate simulation of the daily variation in rainfall depths.

If the relationship between long term annual rainfall and annual runoff is strongly linear (as will be the case when evaporation rates are rarely limited by soil moisture deficits) the random errors in annual BIAS will not propagate into major errors in the prediction of long term runoff. However it should noted that BIAS errors in years where significant soil moisture deficits can potentially build up might lead to potentially large water balance errors in simulated runoff over the full period of record. Evaporation rates will be rarely limited by soil moisture deficits in the wetter catchments in the north and west of the United Kingdom but will occur in most years in drier parts of the country.

A comparison between 61-90 average annual rainfall estimates generated for each catchment using the triangular planes method, with and without additional enhancement through the use of AAR, demonstrated that the 61-90 SAAR for high rainfall catchments is consistently underestimated when AAR is not included within the estimation procedure. This systematic under estimation in high rainfall catchments is attributed to a biased siting of rain gauges. The siting of gauges tends to be biased towards accessible low altitude areas, these are either coastal areas, or areas in which rainfall is generally lower as a consequence of rain shadowing.




**(a) Histogram of percentage differences**




**(b) Percentage difference plotted as a function of mean catchment elevation**

**Figure 2.13 Difference between 40km and 1km resolution MORECS based estimates of 1961-90 average annual PE.**

A scheme for estimating monthly time series of a reference Potential Evaporation demand for any point within the United Kingdom at a 1-km resolution has been developed. This scheme uses monthly MORECS II Penman Monteith PE estimates for short grass. The scheme is based on the interpolation of these estimates to a 1-km grid using empirically derived lapse rates in conjunction with a 1-km resolution elevation grid. This scheme was used to derive a time series of monthly PE data for the catchments used within the study. The differences between 61-90 average annual PE estimates derived using this scheme and those derived directly from MORECS were evaluated for these catchments. This comparison demonstrated that in most of the catchments the catchment PE estimates derived were lower than those derived using the raw MORECS data with a strong positive relationship between the size of the difference and the mean catchment elevation. This is a consequence of the skew within the catchment data set to higher elevation catchments (discussed in Chapter 5). This demonstrates the advantage gained through the use of the 1km interpolation procedure in deriving more realistic PE estimates, particularly for higher elevation catchments.

# 3 Evaluation of rainfall runoff models for use within a regionalisation scheme

## 3.1 MODEL SELECTION

The primary objective of this part of the research was to identify a suitable rainfall runoff model structure for generalisation studies within the United Kingdom. The approach adopted has been to evaluate the utility of selected lumped rainfall runoff models within five case study catchments within East Anglia. Other researchers have used this approach to evaluating models. The World Meteorological Organisation (1974) commissioned the testing of 10 different models on six rivers distributed throughout the world from a flood assessment perspective. Weeks & Hebbert (1980) tested four models on three catchments within South Western Australia. A further study was undertaken by Chiew *et al* (1993) in which six rainfall runoff models were evaluated through their performance in eight Australian catchments while Franchini & Pacciani (1991) evaluated seven well-known models within the Arno basin in Italy.

From a water resources perspective, the basic requirements of a rainfall runoff model are that the model should reliably simulate the processes that control actual evaporation losses and the routing of effective precipitation through the catchment. The low rainfall regime of East Anglia means that summer actual evaporation rates are much lower than the potential evaporation rate. Consequently the application of models within the region represents a good test of the ability of the model loss modules to model the complex relationships between actual evaporation and soil moisture deficit, which in turn are dependent on land use and the soils within a catchment. The small difference between rainfall and actual evaporation within the Anglian region implies that model performance will be extremely sensitive to the ability of a model to accurately model actual evaporation.

A literature review of conceptual and empirical rainfall runoff models suitable for continuous simulation of daily mean flows was undertaken. The review considered the following questions:

- Is there peer reviewed evidence of the use of the models within a generalised context?
- Is there peer reviewed evidence of the model's capabilities?
- Is the model commercially available, public domain or published in full within the literature?
- What are the input data requirements?
- Is the model stochastic, deterministic or hybrid?
- If a model is deterministic, is the model conceptualisation physically based, or empirical in nature, or a mixture of both?
- How complex is the model (a subjective decision made on the basis of the number of parameters and the complexity of the input data requirements)?
- If calibration schemes exist for a model, what optimisation routines and associated objective functions are employed within the schemes?
- If the model is packaged, what analysis functions are available?

This review was not intended to be a definitive review of models, but rather a broad review of approaches to rainfall runoff modelling. The review identified 19 distinct models, these are summarised in Table 3.1. For each model this table presents the author (based on affiliated institution), the processes represented within the model, a subjective measure of complexity based on the number of model parameters (high, medium or low) and a primary reference. From the review the following four models were selected as being broadly representative of the model classes reviewed:

- The Hydrological Simulation Model (HYSIM);
- The Climate, Land-use and Abstraction Model (CLAM) implementation of the Thames Catchment Model (TCM));
- The Probability Distributed Model (PDM);
- Identification of unit Hydrograph And Component flows from Rainfall, Evaporation and Stream flow data (IHACRES).

As discussed in Chapter 1 all of these models, or components of these models have been used to some extent for estimating stream flow within ungauged catchments. The key features of these models are summarised in the following sub-sections

## Table 3.1 Summary of the models reviewed

| Model | Author | Conceptual processes | Com-plexity | Reference |
|---|---|---|---|---|
| Hydrological Rainfall RunOff Model (HYRROM) | Institute of Hydrology | IS, SM, DR,GW | LOW | Blackie & Eeles, 1985 |
| Probability Distributed Model (PDM) | Institute of Hydrology | SS, PDSM, DR,GW | LOW | Moore, 1985 |
| ARNO | Inst Hyd Con Univ Bologna, IT | SS, PE, PDSM, DR, GW | MED | Todini, 1996 |
| Hydrological, Simulation Model HYSIM | R E Manley, Cambridge | SS, IS, DR, MLSM, MLGW, CR | MED | Manley, 1978 |
| Thames Catchment Model | B Greenfield, Thames EA | N zones MLSM, GW | LOW | Greenfield, 1984, NRA R&D Note 268 |
| TANK Model | M Sugawara, Tokyo | TF | LOW | Sugawara, 1995 |
| UBC | Univ Brit Colombia CA | SS, SM, DR, MLGW, | HIGH | Quick, 1995 |
| Precipitation – Runoff Modelling System (PRMS) | USGS-Wat Resources Div | IS, SM, PE, DR, GW | HIGH | Leavesly & Stannard, 1995 |
| Sacramento Catchment Model | US-Dept Of Commerce Nat Weather Service | PDSM, DR, MLGW | HIGH | Burnash, 1995 |
| Streamflow Synthesis and Reservoir Regulation (SSARR) | Hydrologic Engineering Center -US Army Corp | SS, IS, SM, DR, GW | MED | Speers 1995 |
| The HBV Model | Swedish Met & Hydrol Institute | SS, SM, DR, GW | LOW | Bergstrom & Forsam 1973 |
| NAM (inc in MIKE II) | Marketed by DHI | SS, SM, DR, GW | LOW | Nielson & Hansen, 1973 |
| IH-ACRES | Institute of Hydrology / CRES-Australia | ELM, TF | LOW | Littlewood & Parker, 1997 |
| Great Ouse Resource Model (GORM) | WRc | DR, MLGW | LOW | WRc , 1990 |
| SFB | Boughton | SS, GW | LOW | Boughton, 1984 |
| MODHYDROLOG | Univ Melbourne | IS, SM, DR, GW | MED | Chiew & McMahon, 1991 |
| STANFORD IV | NOAA, NWS | IS, MLSM, DR, GW | HIGH | Crawford & Linsley, 1966 |
| XINANJIANG | East China College of Hydraulic Engineering | PDSM, DR, GW | HIGH | Zhao *et al*, 1980 |
| CREC | N/A | SM, DR, GW | LOW | Servat & Dezetter, 1991 |
| VTI-HYMAS | Rhodes University | IS, MLSM, DR, GW | HIGH | Hughes& Sami, 1994 |
| GR3 | CEMAGREF, France | ELM, DR, GW | LOW | Edijatno & Michel, 1989 |

Key for "Conceptual processes" column, Table 3 1

| | | | |
|---|---|---|---|
| PE | = Potential Evaporation | DR | = Direct Runoff (interflow and overland flow, lumped or explicit) |
| IS | = Interception Storages | | |
| SM | = Soil Moisture store | GW | = Ground Water store |
| PDSM | = Probability Distributed Soil Moisture Store | MLGW | = Multi-Layer Ground Water store |
| | | CR | = Channel Routing |
| MLSM | = Multi-Layer Soil Moisture store | TF | = Transfer Function |
| | | ELM | = Empirical Loss Module |

### 3.1.1 The Hydrological Simulation Model (HYSIM)

HYSIM (Manley, 1977,1978,1992a, 1992b) is a complex conceptual model in which the response of the conceptual representation of the hydrological processes is controlled by parameters, many of which the author has sought to relate to physical catchment properties. HYSIM is a seven store conceptual model coupled to a simple hydraulic routing model. This structure is summarised in Figure 3.1. When developing the model the author had the stated primary requirement that the parameters of the model should be physically significant (Manley, 1978). This summary is based on the published information for the model.

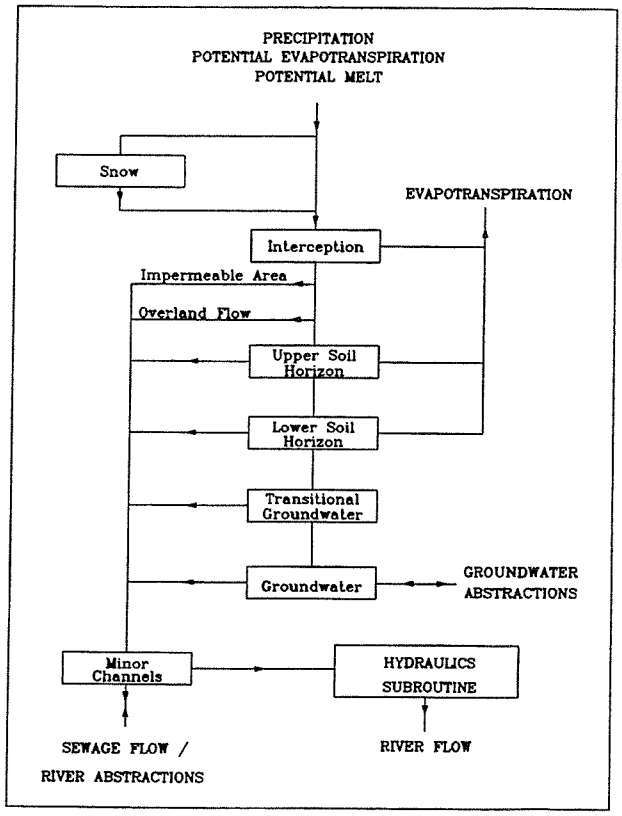



**Figure 3.1 The HYSIM model structure** ***(Source: Manley, 1992a).***

The Snow and interception stores

Precipitation in the form of snow (as defined in the input data) enters a semi - infinite store. If there is snow in storage within time step the outflow is equal to the input melt rate within the time step. The interception store represents detention of water on vegetation. The

maximum depth of the interception store is a calibration parameter. The store receives water from precipitation and snow melt (if any) and loses water by evaporation at the potential rate. Excess precipitation (EP) is partitioned between the upper soil horizon and minor channel storage according to the fractional extent of the catchment defined as being impermeable.

Soil Moisture Store

The soil moisture store consists of two stores, the Upper Soil Horizon (USH) and the Lower Soil Horizon (LSH). The USH represents moisture held in the topsoil (the A soil horizon) whilst the LSH represents moisture below the USH but still within the rooting depth (the B and C soil horizons). The maximum rate at which the store can accept EP is determined by an approximation to Philip's infiltration equation (Philip, 1957). This relationship facilitates the calculation of the maximum, or potential, infiltration rate across the time step. EP routed to the USH in excess of this limiting rate is routed to the minor channels store as overland flow. Evaporation takes place from the USH at the potential rate (minus any loss from the interception storage) if the capillary suction, P, calculated by the model is less than 15 atmospheres. If P is greater than 15 atmospheres then evaporation takes place at a rate reduced in proportion to the remaining depth of water in storage.

The next transfer of moisture is via inter-flow laterally through the USH. The conceptual representation of inter-flow is based on the Brookes and Corey (1971) empirical model for the effective permeability of porous media. The final transfer of moisture is by percolation from the USH to the LSH, where percolation is estimated in an analogous way to inter flow using a non-linear relationship relating percolation to effective saturation. The change in storage within the USH is estimated by combining the equations for infiltration, inter flow and percolation with the continuity equation.

The percolation from the USH forms the input to the LSH. The LSH is configured in a similar way to the USH where the infiltration of percolation is controlled by the ability of the LSH to accept percolation from the USH. Percolation in excess of the infiltration capacity is routed to the minor channels store. Loss from the LSH through inter-flow and percolation to the groundwater is controlled by similar equations to the USH. Evaporation potential that is not met by the USH is met from the LSH, subject to the same suction

pressure constraint that operated in the USH.

The Groundwater reservoirs

The groundwater store is subdivided into two infinite linear reservoirs (Horton, 1938) called the transitional groundwater and deep groundwater stores. The transitional groundwater store which receives percolation from the LSH, is taken to represent the first stage of groundwater storage where direct discharge to surface waters may occur via fissure flow, etc. The outflow from the transitional groundwater store is partitioned between the minor channels store and the deep groundwater store. The deep groundwater store discharges to the minor channels.

The Minor Channels Store and Hydraulic Routing

The minor channels store conceptually represents the routing of flows in minor streams, ditches and, if the catchment is saturated, ephemeral streams. This store uses a triangular Instantaneous Unit Hydrograph (IUH), with the time base equal to 2.5 times the time to peak. The time to peak is estimated using the Flood Studies Report event model equation (Natural Environment Research Council, 1975). The "main" river within HYSIM is represented as a number of hydraulically homogenous reaches.

Velocity of water along a reach is described by the kinematic wave approximation to the Saint Venant equations (Lighthill and Witham, 1955). In this approximation, the wave velocity is the ratio of the incremental changes in flow and hydraulic cross-sectional area along the reach. An empirical model is used within HYSIM for estimating cross-sectional area.

Depending on the configuration, HYSIM has 22 hydrological parameters and six hydraulic routing parameters.

### 3.1.2 The Thames Catchment Model (TCM))

The TCM, which was originally developed by Greenfield (1984), is a conceptual model, based on a simple Penman drying curve based loss module coupled with a series combination of a linear reservoir and a quadratic reservoir. The model is used operationally

within the Thames basin (Moore *et al*, 1989, 1994) and has been used to model the relative impact of weather, land use and groundwater abstraction on low flows in case study catchments across England and Wales Wilby *et al* (1994a &b). This summary of the model is based on the published work of Wilby. The version of the model used for the study was the PC version developed by Wilby.

The structure of the TCM hydrological model is based on the subdivision of a basin into different response zones representing, for example, runoff from aquifer, clay, riparian and paved areas. The zones share the same model structure but have different, appropriate parameter sets. The zonal flows are combined to yield the total catchment runoff. A response zone may be considered to represent a combination of sub-areas within a catchment having similar hydrological characteristics. The conceptual representation of a hydrological response zone in the TCM is illustrated in Figure 3.2. Each zone consists of a two-stage soil moisture store, a linear reservoir, and a quadratic non-linear reservoir connected in series.

The soil moisture store

Within a given zone, water movement in the soil is controlled by the Penman storage model (Penman, 1949) in which a near-surface storage, of depth equal to the rooting depth of the associated vegetation (the root constant depth), drains only when full into a lower storage of infinite capacity. Evaporation occurs at the potential rate (P.E), whilst the upper store contains water and at a lower, actual rate (A.E), when only water from the lower store is available. The threshold deficit at which this lower rate evaporation is initiated is optimised though calibration. The A.E rate is set to 0.3P.E rather than 0.08P.E, as in the original Penman model based on the work of Hyoms (1980) during the 1976 drought. The upper Penman soil moisture store is replenished by rainfall, but a fraction called direct percolation (typically 0.15) is bypassed to contribute directly as percolation to the linear reservoir which may be conceptualised as unsaturated storage. Percolation occurs from the upper Penman store only when the total soil moisture deficit has been made up.

The linear reservoir and non linear reservoirs

The outflow from this reservoir is proportional to the water held within the store. This outflow acts as the input to the non-linear reservoir. The constant of proportionality, the

time constant, controls the response of this reservoir. A quadratic storage function is used to represent the response of the saturated zone. The inflow into this storage is the outflow from the linear reservoir. The outflow from this reservoir is proportional to the square of the volume in storage. To obtain a volumetric flow rate it is necessary to multiply the outflow from the non-linear reservoir by the area of the zone being considered. Each zone within the TCM has nine parameters

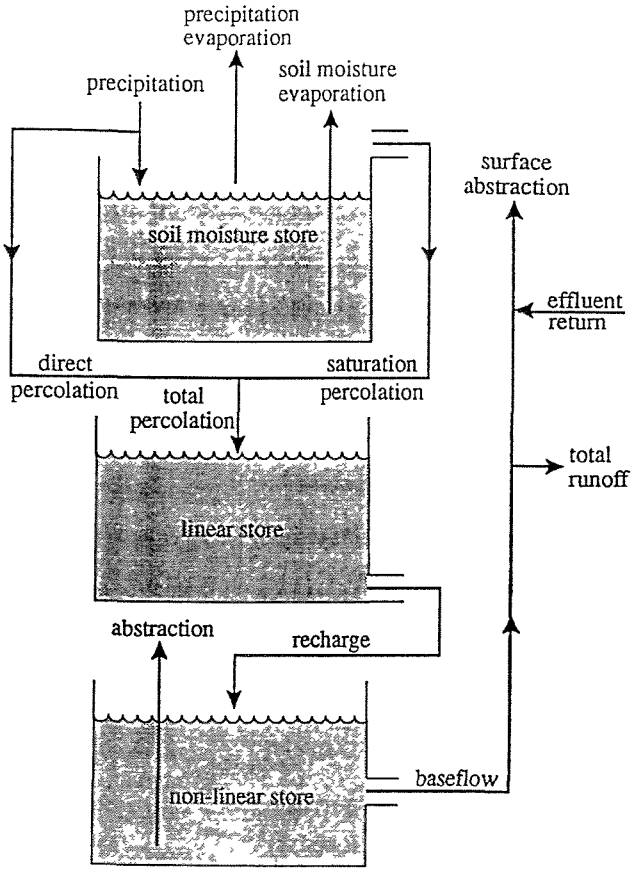



**Figure 3.2 The Structure of the Thames Catchment Model** ***(Source: Wilby et al. 1994a).***

### 3.1.3 The Probability Distributed Model (PDM)

Moore (1985) developed the Probability Distributed Model (PDM). In the context of this evaluation, a fixed configuration was used which represented the most common configuration of the model (Moore *et al*, 1994). The PDM uses a probability-distributed approach for modelling soil storage capacity. The general form of the model is illustrated in Figure 3.3. This summary of the model is based upon the published works of Moore.

Runoff production at a point in the catchment is controlled by the absorption capacity of the soil to take up water. Instead of conceptualising this as a simple store with a given storage capacity, the differing point storage capacities and that the spatial variation of capacity is described by a probability distribution. The most commonly used distribution is the Pareto, or reflected power, distribution. Based on this it is possible to formulate a simple runoff production model which integrates the point runoffs to yield the catchment surface runoff into surface storage. Groundwater recharge from the soil moisture store passes into sub-surface storage. The outflow from surface and sub-surface storages (or reservoirs) forms the model output. This probability-distributed approach to soil storage capacity has also been used by other researchers notably within the ARNO, XINANJIANG models (referenced in Table 3.1), the grid based VIC model, Arnell implementation of the PDM and Jolley and Wheater grid based implementation of the TCM soil moisture store. These grid-based models are discussed in Chapter 1.

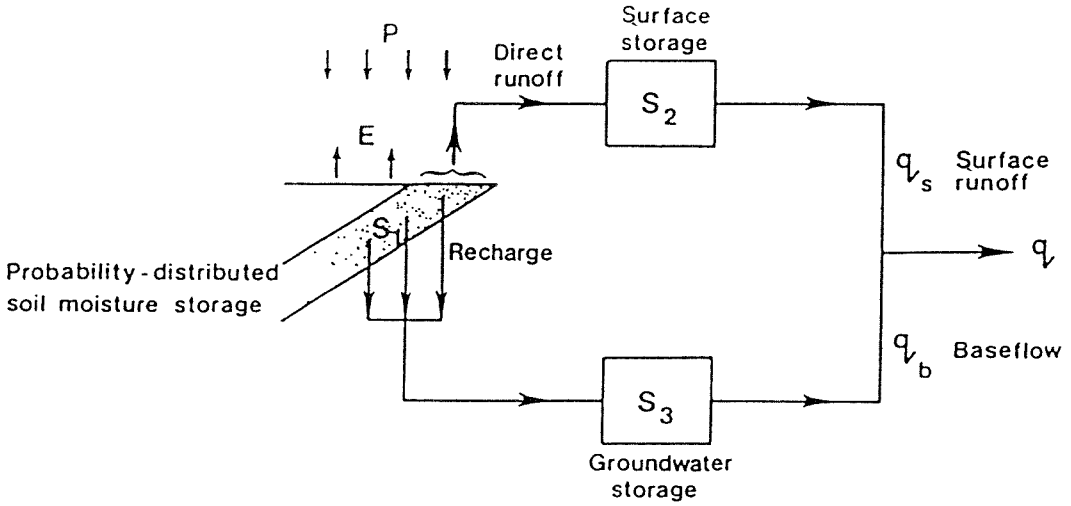



**Figure 3.3 The Structure of the PDM rainfall-runoff model** ***(Source: Moore, 1985).***

As the PDM was finally selected to form the basis of the rainfall runoff model it is described briefly here with the probability distributed concepts discussed more thoroughly in Chapter 4, in the context of model development for this study. The discussion in Chapter 4 also draws from Moore *et al* (1994). The configuration of the PDM used in the evaluation was a Pareto based soil moisture store with the surface runoff routed through two linear reservoirs in series with identical time constants. That part of the effective rainfall attributed to base flow was routed through a cubic non-linear reservoir. Two

options for determining the partitioning of effective rainfall between surface and groundwater flow paths were tested. The first was based on a simple fixed split of runoff from the soil moisture store, with the split parameter optimised during calibration. In the second option a drainage term was included in which the drainage to the slow flow reservoir was inversely proportional to the soil moisture deficit. This configuration is similar to Arnell's grid based implementation. The constant of proportionality was optimised during calibration. This configuration of the PDM has nine parameters.

### 3.1.4 Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Stream flow data (IHACRES)

Overview

The version of IHACRES selected for this study is the PC implementation of the model, PC-IHACRES, V1.0 (Littlewood and Parker, 1997). The first published account of the IHACRES methodology and it's application to two small research catchments in Wales is by Jakeman *et al*, (1990). IHACRES comprises a non-linear loss module in series with either a single linear unit hydrograph (UH) model or, alternatively, two linear unit hydrograph models in parallel or series. This summary of IHACRES is based on the information published by Littlewood & Parker (1997), Jakeman *et al* (1990), Jakeman & Hornberger (1993) and Littlewood & Jakeman (1994).

The input data requirements are restricted to time series of rainfall, stream flow, and temperature. The latter is used within the model to approximate evaporation. The model comprises two modules, in series, as shown in Figure 3.4

Within IHACRES an assumption is made that there is a linear relationship between effective rainfall and stream flow. This allows the application of unit hydrograph theory in which the catchment is represented as a configuration of linear reservoirs acting in series and/or parallel. All of the non-linearity commonly observed between rainfall and stream flow is accommodated in the loss model. Although this does not purport to conceptualise the physical relationships between soil moisture, evaporation and drainage a comparison may be made with the soil moisture stores considered within the preceding models.

Conceptualisation of spatially distributed processes in both the non-linear and linear modules of the IHACRES model is restricted. An advantage of the approach, however, is that the model requires only a small number of parameters. In the typical configuration of the non-linear loss module in series with two parallel linear modules there are three parameters in the non-linear loss module and another three in the linear module, making a total of six parameters overall.

The non-linear (loss) module

The loss module, which estimates effective rainfall, accounts for all of the non-linearity in the catchment-scale rainfall-runoff process. The underlying conceptualisation in this part of the model is that catchment wetness varies with recent past rainfall and actual evaporation. A catchment wetness index is computed at each time step on the basis of recent rainfall and, usually, temperature. The catchment wetness index reflects that a catchment that is already wet will generate more effective precipitation than if it is previously dry. The percentage of rainfall which becomes effective rainfall in any time step varies linearly between 0% and 100% as the index varies between zero and unity. If input data to the loss model are restricted to those for rainfall the catchment wetness is calculated as

$$s_k = Cr_k + (1 - 1/(\tau_w))s_{k-1} \quad s_0 = 0, \qquad \textbf{(3.1)}$$

where $r_k$ is the rainfall depth within the time step and $\tau_w$ is the time constant, or inversely, the rate at which the catchment wetness declines in the absence of rainfall. This time constant is termed the catchment drying constant. A larger value of $\tau_w$ gives more weight to the effect of antecedent rainfall on catchment wetness than a smaller one. The parameter C is a constant of proportionality optimised during calibration so that the volume of excess rainfall is equal to the total stream flow volume over the calibration period, after adjustment for the change in catchment storage between the beginning and the end of the period. The excess or effective rainfall within the time step is calculated from the product of the rainfall and the catchment wetness index. To account for fluctuations in evaporation the catchment drying constant can be modulated by a function that relates this to the third variable, usually temperature.

The linear (UH) module

A full description of the linear hydrograph module is given within the literature. The module is essentially an extension of the simple discrete-time hydrograph such that unit effective rainfall over one data time step produces stream flow $b$ (<1) over the same time step. In each subsequent time step, stream flow is a fixed proportion ($a < 1$) of what it was in the previous time step and thus the flow decays exponentially (at a rate determined by a). The area under the UH (volume of flow) is given by the sum of the infinite geometric series ($b + ab + a^2b + a^3b + ...$) and, by definition, this is one unit. With $0 < a < 1$, this infinite geometric series sums to $b/(1 - a)$. The shape of the UH is completely defined, therefore, by one parameter (either $a$ or $b$). Experience has demonstrated that two UHs in parallel is the optimal configuration identifiable from the input data (Littlewood, *pers. comms*).

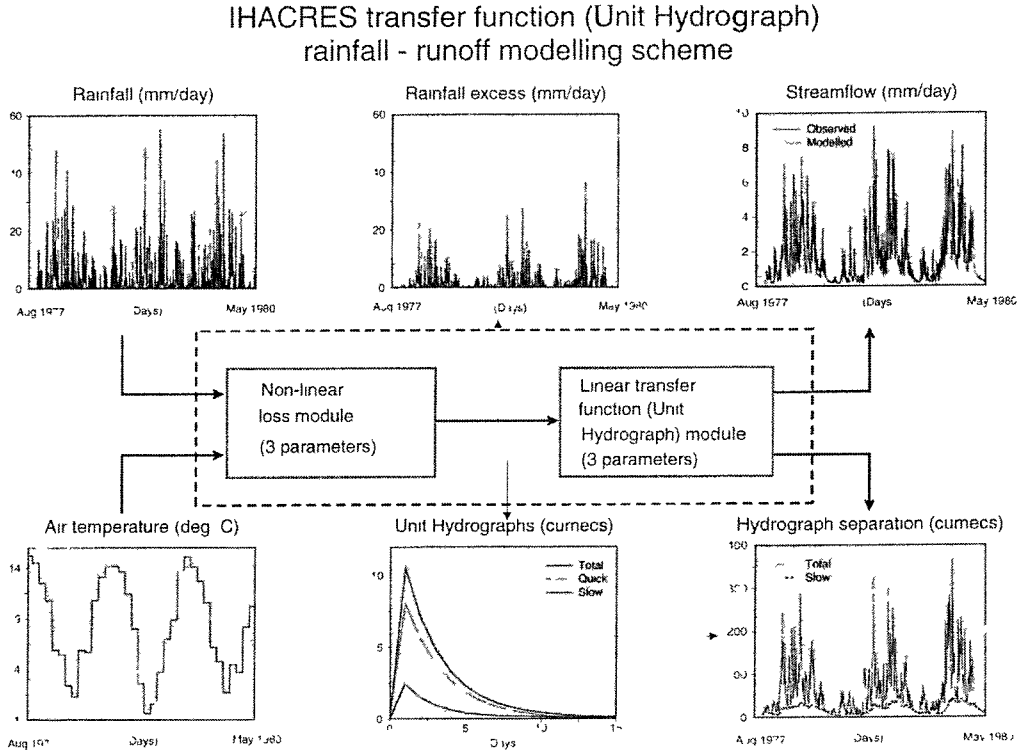



**Figure 3.4 Structure of the IHACRES modelling methodology *(Source: Parker and Littlewood, 1997).***

One of these corresponds to a quick flow component and the other to a slow flow component. The separate UHs sum to give a UH for total stream flow which has a mixed-exponential decay. The response of this second order transfer function can be controlled by any three of the parameters for the two linear reservoirs. A central component of the

IHACRES methodology is the use of the Simple Refined Instrumental Variable (SRIV) technique to estimate the parameters of the linear module for a given set of stream flow data and effective rainfall data received as output from the non-linear module. The detail of the SRIV technique is beyond the scope of this study, however the reader is referred to the work of Jakeman *et al* (1990) for further information.

## 3.2 APPLICATION OF THE MODELS WITHIN THE CASE STUDY CATCHMENTS

### 3.2.1 The case study catchments

The case study catchments selected for the evaluation were the:

- Babingley Brook above the Castle Rising gauging station ( IH Gauge No.33054);
- Sapiston Brook above the Rectory Bridge gauging station (IH Gauge No.33013);
- River Nene above the Orton gauging station (IH Gauge No.32001);
- River Blackwater above Appleford Bridge (IH Gauge No.37010);
- River Box above Polstead Bridge (IH Gauge No.36003).

These catchments were selected to represent a broad cross section of catchment types across the Anglian region and are described in more detail below. The objective of the exercise was to apply the models to the 22-year period of record between 1970 and the end of 1992, where available. The selection of the period of record was restricted by the availability of naturalised flow data for the Blackwater and the Nene. The three-year period 1986-1988 was used as a calibration period (as the flow variability within this period is broadly representative of that across the full 22-year period) and the period either side of the calibration period used for model evaluation. The periods selected for the Babingley Brook and the Sapiston were respectively 1976-1992 and 1970-1990. The Castle Rising gauging station, which replaced an unreliable upstream gauge, became operational in 1976 whilst significant utilisation of the groundwater resources within the Sapiston catchment commenced post 1990.

The Babingley Brook above the Castle Rising gauging station (33054)

The Babingley Brook above Castle Rising has a topographic catchment area of 47.7 $km^2$, however the mean groundwater catchment area is believed to be approximately 86 $km^2$. The gauging station is a triangular profile flow V crump weir and was assigned an A Grade for hydrometric quality at low flows by the Institute of Hydrology (Gustard *et al*, 1992).

The catchment average value of the Meteorological Office Standard Period 1961-90 Average Annual Rainfall (SAAR) is 670 mm/year and the catchment has a gauged runoff of 150-200 mm/yr. The catchment is predominantly unconfined Chalk and thus the flow regime is heavily dominated by groundwater discharge. The land use within the catchment is primarily arable.

The artificial influences within the catchment are dominated by the utilisation of groundwater for public water supply, which constitute 98% of the licenced abstractions. The abstraction time series over the period 1976-1992 have been accumulated at a monthly resolution as part of a naturalisation study undertaken by the Anglian Region of the Environment Agency in 1992 (Watts, 1994). For the model simulations the monthly influence series was partitioned to generate a daily series. The daily flow series at the gauging station were subsequently naturalised by adding in the abstraction time series. This was considered to be appropriate as the significant boreholes are close to the river channel and the abstraction time series has little seasonal variation.

The Sapiston Brook above the Rectory Bridge gauging station (33013)

The Sapiston at Rectory Bridge is a rectangular thin-plate weir gauging an upstream catchment of 206 $km^2$ and was assigned an IH grade A for hydrometric quality at low flows. The catchment SAAR(61-90) is 590 mm/year and the gauged runoff is approximately 105 mm/year. Prior to 1990 the catchment was essentially natural with only minor abstraction for public water supply and agriculture. The catchment is agricultural in nature with geology dominated by Chalk with Boulder Clay cover.

The Nene above the Orton gauging station (32001)

The flow record for the Nene at Orton is a composite record. Flows below 17 $m^3s^{-1}$ are measured at Orton . Flows above 17 $m^3s^{-1}$ are derived by re-scaling flows measured at Wansford, which lies some 12-km upstream from Orton. The structure at Orton consists of a series of sluices, weirs and a lock. The station was assigned a B grade for hydrometric quality under the IH grading system. The station is the lowest on the Nene and gauges an upstream catchment area of 1634$km^2$. The SAAR(61-90) across the catchment is 616 mm/year and the catchment has a gauged runoff of 180-190 mm/yr. The catchment is mainly clay and rural in nature.

The flow record is heavily artificially influenced by direct and indirect abstractions for public water supply, agricultural abstraction and effluent returns. The system is complicated by the abstraction at Wansford for Rutland Water, which is used to supply towns within the catchment, such as Northampton that discharge back into the Nene. In 1992 the Anglian Region of the Environment Agency undertook a programme of naturalising the record flow through decomposition (Fawthrop, 1992). The resultant naturalised flows were used for this study.

The River Blackwater above Appleford Bridge (37010)

The Blackwater above Appleford Bridge is a very rural catchment with a catchment area of 247.3 $km^2$. The gauging structure is a double throated trapezoidal flume assigned an A grade for hydrometric quality at low flows. The catchment SAAR(61-90) is 572 mm/year and the catchment has a gauged runoff of approximately 160mm/yr. The hydrogeology of the catchment is principally Boulder clay over London Clay with Chalk in the headwaters.

The majority of artificial influences on the flow record are associated with small abstractions for agricultural, public water supply and industrial purposes and small sewage treatment works discharges. The primary influence is water transferred from the Stour to the Blackwater as part of the Ely Ouse Transfer Scheme. The transferred water is discharged into the River Pant at Great Sampford in the headwaters of the river. During the 1970-1992 period considered within this study extensive transfers have been made in 1973/4, 1976 and 1989-1992. Small but significant transfers have also been made during 1980, 1984,1986 and during testing in 1971. The gauged flow record was naturalised by

Young & Sekulin (1996). This naturalised record was used for the current study.

The River Box above Polstead Bridge (36003).

The Box above Polstead is rural, natural catchment with a catchment area of 53.9 $km^2$. The gauging structure is a trapezoidal flume with a high flow rated spillway that rarely drowns. The structure was assigned an IH A grade for hydrometric quality. The catchment SAAR(61-90) is approximately 566mm/yr with a gauged runoff of 130 mm/yr. London Clay with Chalk dominates the catchment hydrogeology in the north, all overlain by superficial deposits. The minor artificial influences on the flow record are mainly associated with abstractions for agricultural purposes and sewage treatment plant discharges. The flow record was naturalised for the influence of these minor influences by Young & Sekulin (1996). This naturalised flow record was used for this study.

### 3.2.2 The derivation of input climate data

A catchment, average daily rainfall time series was generated for all catchments using the method of triangular planes with normalisation by AAR as described and evaluated in Chapter 2. The PDM, HYSIM and TCM models all require a time series of catchment average potential evaporation as input to the model, whereas the PC version of IHACRES requires catchment average temperature time series data. This evaluation of suitable rainfall runoff models preceded the development of the national MORECS based PE estimation method described in Chapter 2. For this evaluation MORECS II weekly PE estimates for short grass and temperature estimates were utilised. As discussed in Chapter 2 these data are available at a grid resolution of 40km. Where a catchment intersected more than one MORECS grid cell an area weighted average of cell values was taken. The resultant weekly time series were partitioned to give daily time series for input into the models. This method of generating PE estimates does not take into account the spatial interpolation and altitude corrections described in Chapter 2. However given the low relief of East Anglia this was not thought to be an issue.

### 3.2.3 Application of the models within the case study catchments

The evaluation of the model results is presented in Section 3.3. This section summarises the mode of application of each model.

HYSIM

In all catchments, the objective function used from the model package was the Extremes Error of Estimate (Manley 1992b) (based on the product of the explained and unexplained variance normalised by the product of the simulated and observed stream flow summed over all time steps) which is recommended for use as a general objective function. The recommended procedure for fitting the model, as discussed in the user guide, was adopted. Firstly default values were set for all parameters using the guidance given in the reference manual. The second step was to optimise a potential evaporation correction factor using the Newton-Raphson single parameter optimisation option to ensure mass is conserved over the calibration period. The third step was to use the Rosenbrock search algorithm (Rosenbrock, 1960), in conjunction with visual inspection and manual intervention, to optimise the remaining parameters in the model.

IHACRES

The approach for calibrating IHACRES is based around incremental searching through the parameter space of the loss model and the subsequent solving of the linear, routing model using the SRIV technique. The fit of the model is assessed through visual examination of the modelled flows, the coefficient of determination between observed and simulated flows and the uncertainty associated with the parameter values for the linear module as measured by the Average Relative Percentage Error statistic (ARPE). In the calibration procedure conservation of mass is ensured through the inclusion of a volume-forcing coefficient in the loss module. For this study the full loss module was employed; in this the time constant, $\tau_w$, within a time step is modulated according to a temperature dependent function. In this configuration the response of the loss module is controlled by the volume forcing coefficient, C, the time constant, $\tau_w$, and $f$, the modulation constant which determines how sensitive the modulation function is to temperature. A further term to be considered is a pure time delay between the non-linear and linear modules. The model was calibrated over the period from October 1985 to October 1989. The manual recommends

starting the simulation in October when runoff is generally low to minimise the error in estimating the volume-forcing coefficient. The approach adopted for searching the parameter space in the loss module was to set the time delay to zero and search the parameter space defined by $T_w$ and $f$ for both the first and second order configurations of the linear module. Following the selection of an optimal pairing of $T_w$ and $f$ further simulations were undertaken to optimise the time delay. In all catchments the first order configuration was the optimal one. Where a viable second order solution was obtained the high associated ARPE values indicated that the additional complexity was not warranted.

The Thames Catchment Model

The CLAM implementation of the Thames catchment model was the most problematical model to apply. The primary reasons for this is that it is a complicated model when more than one zone is used. This, coupled with the lack of an interactive or incremental parameter search facility, makes it very difficult to apply. The other consideration is that it was very difficult to assess the model fit using the evaluation statistics available within the package (BIAS (error at mean flow) and Nash-Sutcliffe efficiency (Nash Sutcliffe, 1970)) when more than one zone was used as the statistics were applied to individual zones. Visual inspection of the observed and simulated hydrographs was therefore the major tool used to judge the goodness of fit. The strategy adopted was to set up a zone to model the slow flow component of the hydrograph coupled with a second, quick response zone to capture the residual variability.

Probability Distributed Model

The most common version of the PDM, summarised in Section 3.1.3, was applied within the catchments. In practice it was found that either the direct split or the soil moisture based configurations gave the best results, as judged by visual inspection of the hydrographs and the value of the sum of squares objective function available within the PDM package. The direct split option was used within the Babingley and Box catchments. The calibration strategy was to use the automatic calibration facility in conjunction with manual intervention to obtain a best fit based on the value of the objective function, the reasonableness of parameter values that have a clear physical correspondence and visual inspection of the hydrograph.

## 3.3 EVALUATION OF MODEL PERFORMANCE WITHIN THE CASE STUDY CATCHMENTS

The objective of the evaluation exercise was to look at the performance of the individual models within each catchment and, from this analysis to identify whether any general statements can be made about the relative merits of the four models and their packages. The models were applied using the packaged objective functions and graphical displays. To make comparisons between the models it was necessary to use a set of common goodness of fit tests, these are presented in Section 3.3.1.

The individual catchment assessments undertaken using the goodness of fit tests are presented in Section 3.3.2. A ranking scheme was applied to draw out general statements about model performance across the five catchments. This ranking scheme and the application to case study catchments is presented in Section 3.3.3.

### 3.3.1 Evaluation criteria

All of the packages for the models under evaluation advocated the use of one or more mathematical descriptions, or objective functions. The authors also generally recommend that visual inspection of the hydrograph should form part of the calibration process. This combination of quantitative and qualitative goodness of fit tests represents the classical approach employed when calibrating a rainfall runoff model. A full discussion of quantitative objective functions and qualitative goodness of fit measures is presented in Chapter 6 in the context of fit criteria used in the calibration of the regionalisation rainfall runoff model. For the purposes of evaluating the models the following measures were employed:

- BIAS - expressed as the difference between observed and simulated mean flow presented as a percentage of the observed mean flow;
- $R^2$ -the proportion of observed variance explained by the modelled flows;
- graphical comparison of observed and simulated flow duration curves;
- graphical comparison of observed and simulated hydrographs;
- graphical analysis of summary statistics for the observed flow duration curve.

For the last measure the flow duration curve was derived for the observed flow time series by ranking in order of size and calculating an exceedence percentile for each flow, whilst retaining the date associated with each flow. This is equivalent to assigning a flow exceedence percentile to each date. Twelve key percentile points were considered. For each percentile point, the observed flow data and associated dates falling within the data range of ±0.5% around the point were extracted. For each extracted date the corresponding flow was extracted from the simulated time series. This selection process yields N/100 simulated and observed pairs for each exceedence percentile, where N is the total number of data points in the period being considered.

The performance of the model at each percentile point is assessed by calculating the BIAS (the average of the difference between the observed and simulated flows expressed as a percentage of the observed flow) and Coefficient of Variation (CV) across the N/100 pairs at each point. These are then plotted as a function of exceedence percentile. The BIAS plot provides information as to whether the model consistently under or over predicts at particular flows, whilst the CV plot provides information as to the consistency of the model at particular flows, which can be regarded as a measure of model stability at the percentile point. The CV is used to facilitate comparison between different parts of the flow regime. As the number of pairs for each percentile point is much smaller for the calibration period than for the validation period, direct comparison between the results of a model within the calibration period and the validation period should not be made.

### 3.3.2 Evaluation of model performance within the case study catchments

Within the reporting constraints of this thesis it is not possible to review the results for all catchments in detail. The results for the Babingley Brook are presented in this section to illustrate the process. Within each catchment, the models were assessed both within the calibration period and across the modelled period either side of this period, termed the evaluation period. The graphs for the Babingley Brook are presented in Appendix A. This appendix contains graphs for the observed and simulated example hydrographs, flow duration curves, percentile BIAS and percentile CV plots for both the calibration and evaluation periods.

The Babingley Brook at Castle Rising

*Calibration period*

From inspection of the observed and simulated hydrographs for 1988 (Figure A.1) it appears that none of the models simulated the winter storm events well and that general flow recession characteristics through the year are best modelled by IHACRES and the TCM. The rate of recession for the PDM and HYSIM is too low. HYSIM also fails to model the recovery of flows at the end of the year. None of the models seem to model the response to summer storms well. The TCM does not respond at all whilst HYSIM over predicts the response to large summer storms and fails to pick up the smaller events. The recession rates for response to the summer events are too low for both the PDM and IHACRES. All models route the majority, if not all, of the effective rainfall through a single slow response reservoir. This explains the poor response of PDM, IHACRES and the TCM to summer storms. The origin of the behaviour of HYSIM to large summer storms is less clear, although the behaviour may be associated with the conceptualisation of inter flow within the upper and lower soil horizons.

The flow duration curves presented in Figure A.2 and the mean error at percentile points (Figure A.3) demonstrate that the PDM and HYSIM are the closest in simulating the observed distribution of flows. The CV plot (Figure A.3) shows that the PDM consistently has the lowest CV, followed by IHACRES. The TCM has a low CV at low flows and HYSIM has a high CV.

**Table 3.2 Summary statistics for the Babingley Brook**

| | PDM | IHACRES | TCM | HYSIM |
|---|---|---|---|---|
| **Calibration** | | | | |
| Bias | -1.97 | 2.73 | -15.61 | -0.76 |
| $R^2$ | 0.95 | 0.88 | 0.88 | 0.86 |
| **Evaluation** | | | | |
| Bias | -2.11 | 1.58 | -24.91 | -2.46 |
| $R^2$ | 0.93 | 0.89 | 0.66 | 0.88 |

Summary BIAS and $R^2$ statistics are presented within Table 3.2 for model simulations within the calibration and evaluation periods. In the calibration periods the highest $R^2$ values were for the PDM, HYSIM had the lowest BIAS, whilst the TCM had the largest BIAS but has the same $R^2$ value as IHACRES.

*Evaluation period*

The observed and simulated hydrographs are presented for a dry year (1992) in Figure A.4. The flows simulated by the TCM are consistently lower than the observed, with little or no response to either short-term events or the onset of recharge in the September. HYSIM did not simulate the onset of recharge until November and also consistently underestimates the base flow. Once again the "spiky" response to summer storms is observed. The PDM and IHACRES markedly overestimate the flows at the start of the year but correctly pick up the catchment response to recharge. For the majority of the time both the PDM and IHACRES also significantly over estimate the flows, although the overall bias for the evaluation period is low for both models. The flow duration curve plots, Figure A.5, show that the gradient and hence the variance of the flow distribution simulated by the TCM is close to that of the observed, although the simulated flows are consistently lower than the observed. The distribution fits of both IHACRES and HYSIM simulated flows are good, The PDM underestimates the high flows and overestimates the low flows. The mean error plots (Figure A.6) are consistent with the flow duration plots although the CV plots (Figure A.7) demonstrate that the PDM is much more consistent in the predictive error than the other models. HYSIM and IHACRES are broadly similar with respect to consistency whilst the TCM has the largest CV. The summary statistics over the evaluation period show that the PDM has the highest $R^2$ value whilst IHACRES has the lowest BIAS.

### 3.3.3 Inter-catchment and model comparisons

The analysis of the results from the case study catchments (presented in Section 3.3.2 for the Babingley Brook) demonstrated how difficult it is to draw firm conclusions about the performance of the individual models within the case study catchments with goodness of fit tests often providing conflicting, or inconclusive results. However the results of the exercise demonstrated that a reasonable distributional fit (as described by the flow duration curve) maybe obtained when the time series fit may be very poor. This is of concern when

evaluating model performance and as a consequence the flow duration statistics are not included in the comparison of model performance across catchments. A generalised ranking scheme for target goodness of fit tests was developed and used to assess:

- how amenable the flow regimes of the individual catchments were to modelling using simple lumped models;
- if any of the models could be identified as performing more consistently better than others.

The application of the ranking scheme in these contexts is presented below. The goodness of fit test statistics used within the ranking scheme were:

- Bias;
- $R^2$;
- mean percentile error (the average of the dimensionless error for the 5, 10, 15, 20, 30, 50, 70, 80, 90 and 95 exceedence percentiles);
- stability (the average of the CV of the dimensionless error for the 5, 10, 15, 20, 30, 50, 70, 80, 90 and 95 exceedence percentiles).

The latter two test statistics, whilst not statistically rigorous, attempt to numerically summarise the information presented graphically for individual catchments. In the ranking scheme analysis three scenarios were considered the:

- goodness of fit over the calibration period;
- goodness of fit over the evaluation period;
- goodness of fit over the calibration period and the change in goodness of fit between the calibration and evaluation periods.

For the third scenario, the sum of the modulus of the departure from a perfect fit in the calibration period and the difference between the quality of fit in the calibration period and the evaluation period was used to summarise the performance of the individual goodness of fit tests over the two periods.

Inter-catchment comparisons

This comparison exercise was undertaken to assess, relatively, how well the flow regimes of the catchments could be represented by lumped rainfall runoff models. The application of the ranking scheme in this comparison is discussed with respect to one statistic. For each model the goodness of fit was assessed in each catchment according to each of the test statistics and the catchments ranked according to the value of the test statistic. The average rank across the four models was then taken to given an overall catchment rank for each test statistic. An example for the BIAS statistic over the calibration period is shown in Table 3.3. The average ranks for each test statistic were then collated for each scenario. These are presented in Table 3.4.

**Table 3.3 Example ranking of catchments by model for the BIAS statistic**

| | Calibration Period | | | | |
|---|---|---|---|---|---|
| **Catchment** | **PDM** | **IHACRES** | **TCM** | **HYSIM** | **Mean Rank** |
| Babingley | 1 | 3 | 3 | 1 | 1 |
| Sapiston | 3 | 4 | 1 | 3 | 2 |
| Nene | 2 | 5 | 4 | 2 | 3 |
| Blackwater | 4 | 2 | 2 | 5 | 3 |
| Box | 5 | 1 | 5 | 4 | 5 |

The overall picture produced by this ranking scheme shows that for all scenarios the best model fits were obtained for the Babingley Brook and the worst for the River Box. The Blackwater was consistently fourth. The Nene had an over all rank of 2 over the calibration and 3 over the evaluation period. The Sapiston a rank of 3 over the calibration period but has a rank of 2 over the evaluation period. When considering the goodness of fit in the calibration period and the stability of that goodness of fit between the calibration period and the evaluation period (scenario 3) the rank for the Sapiston is 2 compared with 3 for the Nene.

**Table 3.4 Ranking of catchments by model and scenario**

| | **Calibration** | | | | |
|---|---|---|---|---|---|
| | Babingley | Sapiston | Nene | Blackwater | Box |
| Bias | 1 | 2 | 3 | 3 | 5 |
| $R^2$ | 1 | 3 | 2 | 4 | 5 |
| Mean % err. | 1 | 3 | 2 | 4 | 5 |
| Stability | 1 | 2 | 2 | 5 | 4 |
| ***Overall*** | ***1*** | ***3*** | ***2*** | ***4*** | ***5*** |
| | **Evaluation** | | | | |
| | Babingley | Sapiston | Nene | Blackwater | Box |
| Bias | 1 | 4 | 2 | 3 | 5 |
| $R^2$ | 1 | 2 | 3 | 4 | 5 |
| Mean % err. | 1 | 2 | 3 | 4 | 5 |
| Stability | 1 | 2 | 3 | 4 | 5 |
| ***Overall*** | ***1*** | ***2*** | ***3*** | ***4*** | ***5*** |
| | **Combined** | | | | |
| | Babingley | Sapiston | Nene | Blackwater | Box |
| Bias | 1 | 4 | 3 | 2 | 5 |
| $R^2$ | 1 | 2 | 3 | 4 | 5 |
| Mean % err. | 1 | 2 | 3 | 4 | 5 |
| Stability | 1 | 2 | 4 | 3 | 4 |
| ***Overall*** | ***1*** | ***2*** | ***3*** | ***4*** | ***5*** |

Inter-model comparisons

A similar approach to the inter-catchment comparisons was adopted for the inter-model comparisons. For each catchment the goodness of fit was assessed for each model according to each of the test statistics and the models ranked according to the value of the test statistic. The average rank across the five catchments was then taken to given an overall model rank for each test statistic. An example for the BIAS statistic over the calibration period is shown in Table 3.5.

**Table 3.5 Example ranking of models by catchment for the BIAS statistic**

| | | **Calibration Period** | | |
|---|---|---|---|---|
| **Catchment** | **PDM** | **IHACRES** | **TCM** | **HYSIM** |
| Babingley | 2 | 3 | 4 | 1 |
| Sapiston | 4 | 2 | 1 | 3 |
| Nene | 2 | 4 | 3 | 1 |
| Blackwater | 3 | 1 | 2 | 4 |
| Box | 3 | 1 | 4 | 2 |
| **Mean rank** | 3 | 1 | 3 | 1 |

The average ranks for each test statistic were then collated for each scenario. These are presented in Table 3.6. Within the calibration phase the PDM scores the highest overall rank followed by HYSIM, IHACRES and the TCM in that order. With the exception of the BIAS statistics the scorings for individual test statistics is very consistent. The BIAS rankings reflect that both IHACRES and HYSIM include calibration factors to ensure that mass is conserved over the calibration period.

Over the evaluation period, HYSIM scores the highest overall rank followed jointly by IHACRES and the PDM with the TCM scoring the lowest rank. The scorings for individual statistics are consistent with the calibration period for HYSIM and the TCM. The promotion of HYSIM to rank 1 for the mean error and stability indices is a consequence of the degradation of the PDM scores for these indices. IHACRES retains the highest rank for the BIAS statistics demonstrating the utility of calibrating to ensure mass is conserved.

**Table 3.6 Ranking of models by catchment and scenario**

| | **Calibration** | | | |
|---|---|---|---|---|
| | PDM | IHACRES | TCM | HYSIM |
| Bias | 3 | 1 | 4 | 2 |
| $R^2$ | 1 | 3 | 4 | 2 |
| Mean % err. | 1 | 3 | 4 | 2 |
| Stability | 1 | 3 | 3 | 2 |
| overall | 1 | 3 | 4 | 2 |
| | **Evaluation** | | | |
| | PDM | IHACRES | TCM | HYSIM |
| Bias | 3 | 1 | 3 | 2 |
| $R^2$ | 1 | 3 | 4 | 2 |
| Mean % err. | 3 | 2 | 4 | 1 |
| Stability | 2 | 4 | 3 | 1 |
| overall | 2 | 2 | 4 | 1 |
| | Combined | | | |
| | PDM | IHACRES | TCM | HYSIM |
| Bias | 3 | 1 | 4 | 2 |
| $R^2$ | 1 | 3 | 4 | 1 |
| Mean % err. | 2 | 3 | 4 | 1 |
| Stability | 1 | 3 | 3 | 1 |
| overall | 2 | 3 | 4 | 1 |

When looking at the rankings for the combined score, HYSIM scores the highest rank. This is consistent across all statistics, with the exception of BIAS, where IHACRES has the highest combined rank. The PDM scores the second highest rank and, with the exception of BIAS, the PDM is ranked either second or joint first with HYSIM. IHACRES is ranked third and the TCM is ranked fourth. The overall ranks for the models are very consistent with the ranks for the individual test statistics for scenario 3.

## 3.4 SUMMARY OF THE MODEL EVALUATION STUDY

Looking at the calibration and evaluation periods, the best model fits were consistently obtained for the Babingley Brook and the worst for the River Box. The model fits for the Blackwater were consistently fourth. During the calibration periods the model fits were better for the Nene than the Sapiston, however over the evaluation period better model fits were obtained for the Sapiston than for the Nene. The analysis did not identify whether particular models were more suitable than others for specific catchment types.

During the period of record considered, the Sapiston and Box catchment were relatively natural. When the Ely Ouse scheme is not operating, the Blackwater catchment is essentially natural and, given the transient nature of the schemes operation, the errors in the naturalised flow records associated with the Ely Ouse transfer scheme will not have a major impact upon the quality of the flow record. As the hydrometric quality of the flow record is good, it is difficult to see why the performance of the models should be worse in the Blackwater and Box catchments than the other catchments without further investigations.

The Nene is subject to some complex artificial influences and, given the poor data quality associated with the majority of influences and the temporal variability of the quality, it is quite likely that time dependent artifacts of the influences remain within the naturalised flow record. This may account for why, generally, the quality of the model fits were much better in the calibration period than the evaluation period. Recent discussions staff within the Anglian region supports this view. The parameters identified over the calibration period may well be compensating for these errors.

The case study catchments are amongst some of the driest gauged catchments within the United Kingdom. The treatment of evaporation and the modelling of actual evaporative losses are primary issues when modelling these catchments. Modelling in these dry catchments is thus a good test of the performance of the loss modules within rainfall runoff model. However the issue of errors in the input data must not be ignored. In these dry catchments the gauged runoff is in the order of 100-150 mm/yr, the consequence of relatively small errors in the estimation of catchment rainfall and evaporation/temperature may result in quite major errors in the modelled runoff. For example, taking a crude water balance a 5% error in an estimated rainfall of 600mm/yr may result in a water balance error of up to 30% in the gauged runoff. The issues associated with the propagation of error in climatic data within rainfall runoff models were discussed in Chapter 2.

Obviously the model parameters derived during optimisation will tend to compensate for any errors within the input data, including stream flow data. However, this may lead to structural problems within the model which, coupled with the likely random nature of errors in the input data, will reduce the quality of the model fit over the evaluation period, as in the case of the Nene.

It is important to draw the distinction between the model structure and the packaged optimisation procedures and associated objective functions. The performance of the model will be strongly influenced by the choice of objective function and the efficiency of the optimisation scheme will be strongly influenced by how identifiable, or unique model parameters are, which is a function of the model structure. All of the aforementioned will be influenced by input data quality. On the basis of these considerations it is not possible to definitively conclude that one model is better than another model.

From a technical viewpoint, the ranking scheme adopted demonstrated that the PDM was the most consistent model across the calibration period followed by HYSIM, IHACRES and then the TCM. HYSIM gave better results over the evaluation period than the PDM and when jointly considering the performance in the calibration period and the departure from that performance in the evaluation period HYSIM was the most consistent of the four models. The PDM was the second most consistent model overall, followed by IHACRES and the TCM in that order.

On first sight, it is somewhat surprising that the PDM is the most consistent model across the calibration period but not across the evaluation period. This behaviour is related to the fact that the PDM package was ranked third with regard to minimising BIAS over the calibration period. The mean BIAS for the PDM was 12% over the calibration period and 24% over the evaluation period. This contrasts markedly with IHACRES and HYSIM where the BIAS is small both over the calibration and simulation period. This can be attributed to the fact that both of these models formally ensure that mass is conserved over the calibration period as part of the calibration procedure; IHACRES by means of the volume forcing constant and HYSIM by a scaling factor applied to the P.E. estimates. The PDM, in contrast, has no such calibration procedure; the calibration is based on visual interpretation and the value of a least square objective function. Within the PDM the BIAS error in calibration propagates over the evaluation period and consequently impacts upon the other evaluation measures. The measures used in IHACRES and HYSIM for minimising BIAS appear to also be effective over the evaluation period. This highlights the importance of ensuring that mass is conserved during the calibration period.

One issue that arose during the trials was that of parameter covariance. Whilst this was not formally investigated within the model evaluation process this problem was observed during the calibration procedures for all models and was particularly noticeable within HYSIM. HYSIM is a very complex model, and the use of default values for many of the parameters within the model must raise the question of whether this level of complexity is warranted. Furthermore the strong structural interrelationships between the primary parameters must be a cause for concern regarding parameter identifiability if the model is to be calibrated as opposed to being populated with *a priori* parameter estimates. The model evaluation study has highlighted the:

- importance of ensuring mass is conserved by the model;
- limitations of least square based objective functions when calibrating a model for water resource issues;
- need to separate the performance issues relating to the model from those related to the packaging of the model (including calibration schemes).

When considering the performance of the models over a range of objective functions, HYSIM and the PDM were markedly more effective that IHACRES and the TCM. The evaluation study did not address the issues of parameter identifiability or the ability of the model package to identify the global minimum of the objective function space. These issues become more complex as the number of parameters a model has increases.

On these considerations the PDM philosophy of a statistically distributed soil moisture store was selected to form the basis of a rainfall runoff model for UK regionalisation. The development of the regionalisation model and calibration framework is discussed in Chapter 4.

# 4 Rainfall Runoff Model Development

When assessing the resource available within a catchment, it is necessary to be able to quantify both the average and reliable yield from the catchment. The reliable yield will be dependent upon the competing requirements for water and the availability of that water. The availability is normally constrained by the magnitude of low flow events within the catchment and the frequency of the low flow events. If a regionalised model is to be useful for quantifying resource availability at the ungauged river reach, it is essential, therefore, that the model can model catchment daily mean flows, particularly low flows effectively and replicate the mean daily flow. It is not necessary for the model run on a shorter time step than a day or for the model to be able replicate the catchment behaviour at high flows, other than to ensure that mean flow is modelled correctly.

In the context of a model for regionalisation, it is also advantageous to assume a catchment water balance approach. In this approach, it is assumed that all parts of a catchment enclosed by a boundary defining the extent of the catchment above a point can contribute to river flow at the point. In the context of this study, this boundary has been defined as the topographic boundary. This assumption is useful, as the catchment area is then a model parameter that is defined *a priori*. The assumption also ensures that it is possible to identify meaningful extents for catchment climatic data and characteristics. The limitation of the approach is that systematic errors may be introduced through errors in the estimation of the contributing catchment area or violation of the closed water balance at the point in question. Both these problems commonly occur in phreatic groundwater catchments. The closed water balance assumption is flawed in these catchment as groundwater boundaries tend not be static and rarely coincide with the topographic divide. Furthermore, as the river can be regarded as an exposure of the water table in this type of catchment, it is quite common to have a significant bypass of the channel as result of subsurface groundwater flow.

Following on from the model evaluation study presented in Chapter 3, the Probability Distributed Soil-moisture Module (PDSM) component of the PDM model of Moore (1985) was selected as the basis of a model for the regionalisation studies. The PDSM was used in a modelling scheme with a conventional quick and slow flow routing module for

representing the hillslope and ground water routing of effective precipitation from the soil moisture store. Two configurations were used for the PDSM.

- Configuration A (MODA). A treatment of soil moisture behaviour in which a drainage term for the loss module was included. In MODA the drainage from the PDSM was routed through the slow flow reservoir and the outflow from the PDSM routed through the quick flow reservoir. An interception model was included within MODA to provide some provision for rainfall evaporating at the potential rate, even when soil moisture deficits were significant.
- Configuration B (MODB). A simple treatment of soil moisture and evaporation mechanisms was used for this model. The interception losses from different vegetation types were ignored and the division of effective rainfall between the quick and slow flow routing components was based upon a fixed division.

MODB was introduced to address some of the parameter identifiability problems found with MODA. These issues are discussed in detail within Chapter 6. The model structure for the model configurations is presented in Figure 4.1. The models each consist of three modules:

- the distributed soil moisture store;
- an evaporation module;
- a routing module consisting of two linear storage reservoirs.

These three modules are presented in Sections 4.1 to 4.3. with the distinction made between the two model structures in each case.

## 4.1 THE PROBABILITY DISTRIBUTED SOIL-MOISTURE MODEL

The soil moisture store within the PDSM is presented here in the context of the distribution form adopted for the soil moisture store, the treatment of evaporative losses and the partitioning of effective runoff between quick and slow routing paths. Runoff production at a point in the catchment is controlled by the absorption capacity of the soil to take up water. This can be conceptualised as a simple store with a given storage capacity, c'.

Within a time interval, the store receives water from rainfall, P, and loses water by evaporation, E, until either the storage fills and spills, generating direct runoff, q, or empties and ceases to lose water by evaporation. The behaviour of this store is given by

$$q = \begin{cases} P - E - (c' - S_o) & P > c' + E \\ 0 & P \le c' + E \end{cases}, \qquad \textbf{(4.1)}$$

where $S_o$ is the initial depth of water in storage.

In the PDSM it is considered that different points in a catchment have differing storage capacities and that the spatial variation of capacity can be described by a probability distribution. The points differ from each other only with regard to their storage capacity. The storage capacity at any point, c, may then be considered as a random variate with probability density function, f(c), so that the proportion of the river basin with capacities in the range (c, c + dc) will be f(c)dc. In the PDSM it is assumed that all points are inter-connected so that hydraulic gradient between the point stores at any point in time is zero.

The water balance for a catchment, with storage capacities distributed in this way, is constructed as follows. Assuming that the catchment is initially dry at the start of a time interval and receives a rainfall depth P over a time interval. Over the interval the point stores will fill to a depth P unless they are of lesser depth than P, in which case they will fill during the time interval and generate runoff. The frequency of occurrence of a given store depth is given by the probability density function. The actual runoff produced over the catchment must therefore be obtained by weighting the depth produced by a store of a given depth by its frequency of occurrence, as expressed by f(c).

At the end of the time interval stores of depth less than P are generating runoff. Calling the capacity below which all stores are full at some time t the critical capacity, C*, (C*=P in the present example), the proportion of the basin containing stores of capacity less than or equal to C* is

$$\text{prob}(c \le C^*) = F(C^*) = \int_o^{C^*} f(c)dc. \qquad \textbf{(4.2)}$$

The function F(c) is the distribution function of store capacity and is related to the density function, f(c), through the relation f(c) = dF(c)/dc. This proportion is also the proportion of the basin generating runoff, so that the contributing area at time t for a catchment of area A is

$$A_c(t) = F(C^*(t))\,A\,. \qquad \textbf{(4.3)}$$

The direct runoff rate per unit area from the basin is the product of the net rainfall rate, $\pi(t)$, and the proportion of the basin generating runoff, F(C*(t)), after taking into account interception, evaporative and potential drainage losses from the store. This given by

$$q(t) = \pi(t)F(C^*(t))\,. \qquad \textbf{(4.4)}$$

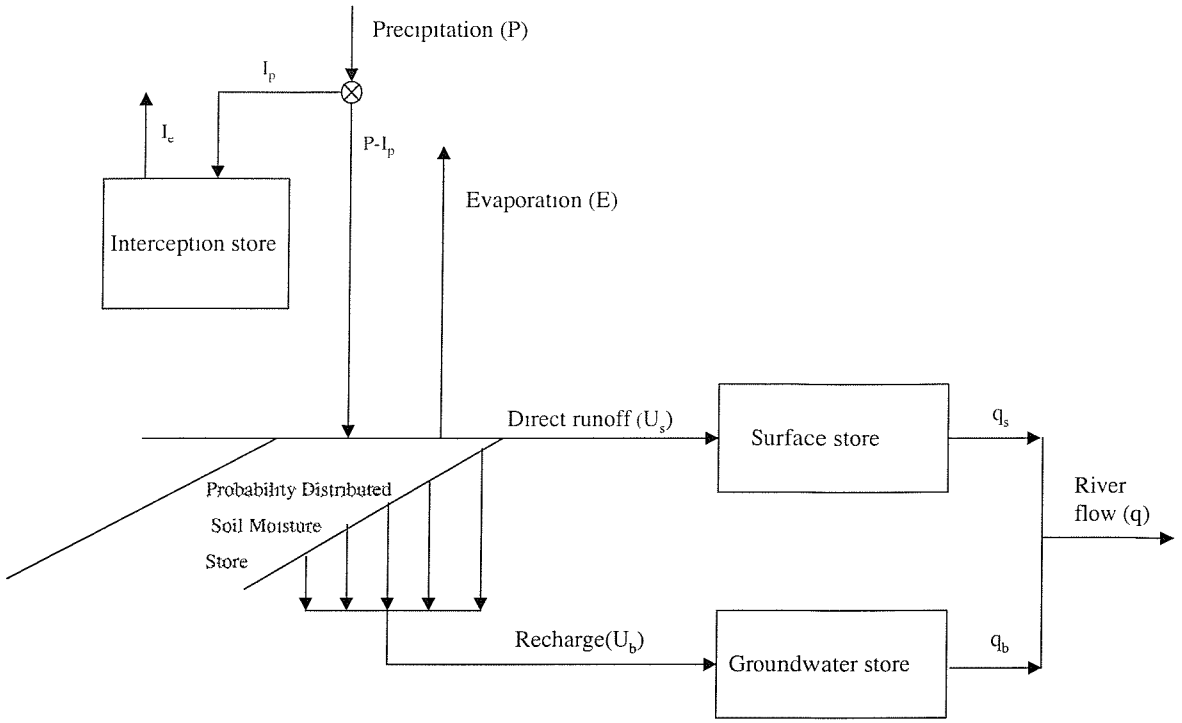
Considering now the i'th wet time interval, (t, t+Δt), in which precipitation, $P_i$, exceeds evaporation, $E_i$, yielding a nett rainfall within the interval of $\pi_i = P_i - E_i$. Then the critical capacity, C*(τ), will increase over the interval according to

$$C^*(\tau) = C^*(t) + \pi_i(\tau - t) \qquad t \le \tau \le t + \Delta t\,, \qquad \textbf{(4.5)}$$

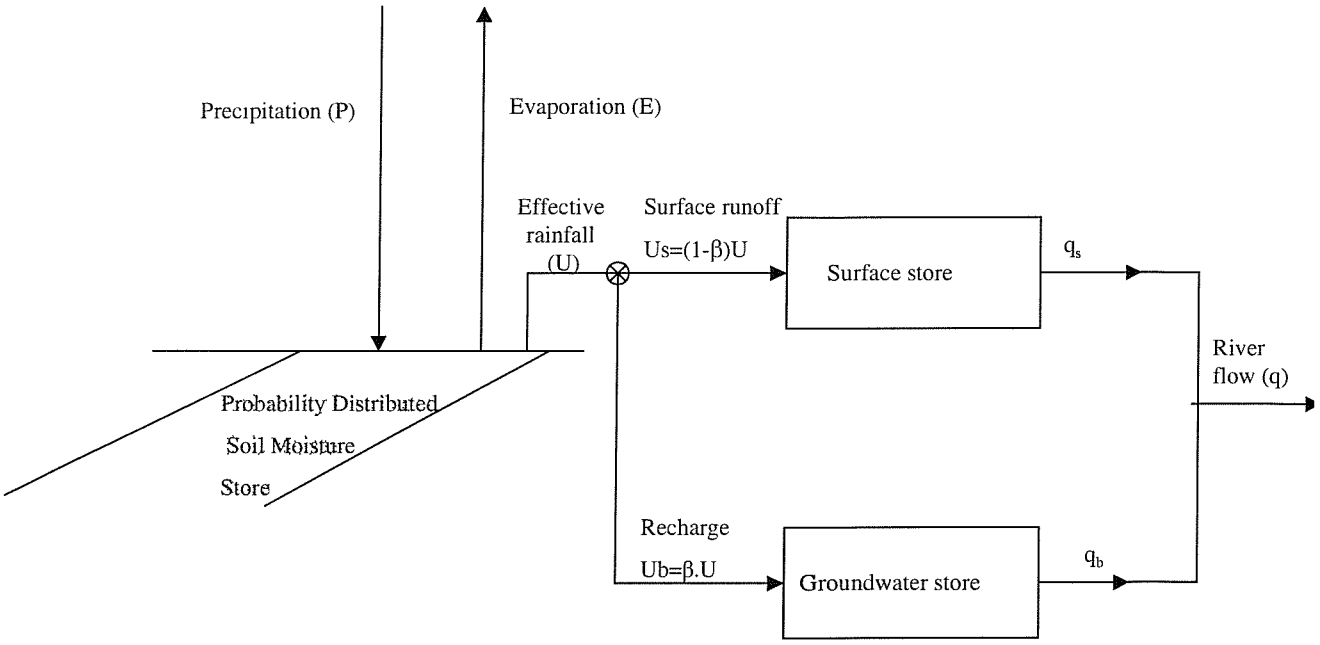
the contributing area will expand according to (4.3), and the volume of basin direct runoff per unit area produced over this interval will be

$$V(t + \ t) = \int_t^{t+\ t} q(\tau)d\tau = \int_{C^*(t)}^{C^*(t+\ t)} F(c)dc\,. \qquad \textbf{(4.6)}$$

During dry periods potential evaporation will deplete the water content of the storages with water moving between stores to equalise the depth of stored water at different points within the basin. Therefore, at any time all stores will have a water content, $C^*$, irrespective of their capacity, unless this is less than $C^*$, when they will be full.




**(a) MODA**




**(b) MODB**

**Figure 4.1 Diagrammatic representation of conceptual model structures.**

Crucial to the PDSM is that a unique relationship exists between the water in storage over the basin as a whole, S(t), and the critical capacity, C (t), and in turn to the instantaneous rate of basin runoff production, Q(t). The total water in storage, at any point in time, is given by the sum of the water held in the proportion of the stores that have a depth less than or equal to the critical capacity and the water held in the proportion of stores which have a capacity greater than the critical capacity, and are not full. This is expressed mathematically as

$$\begin{aligned} S(t) &= \int_0^{C^*(t)} cf(c)dc + C^*(t)\int_{C^*(t)}^{\infty} f(c)dc \\ &= \int_0^{C^*(t)} (1 - F(c))dc . \end{aligned} \qquad \textbf{(4.7)}$$

For a given value of storage, S(t), this can be used to obtain $C^*(t)$ which allows the volume of direct runoff, V(t+Δt), to be calculated using equations (4.6) together with (4.5). The total available storage in the basin, Smax, is given by

$$S_{max} = \int_o^{\infty} cf(c)dc = \int_o^{\infty} \left(1 - F(c)\right)dc = \bar{c}, \qquad \textbf{(4.8)}$$

where $\bar{c}$ is the mean storage capacity over the catchment. During a period when no runoff is generated the soil moisture storage accounting is given by

$$S(\tau) = S(t) + \pi_1(\tau - t) \qquad t \le \tau \le t + \Delta t,\ 0 \le S(\tau) \le S_{max} . \qquad \textbf{(4.9)}$$

When runoff generation does occur then the volume of runoff produced, V(t+Δt), is obtained using (4.6), and then continuity gives the final storage as

$$S(t + \Delta t) = \begin{cases} S(t) + \pi_1 \Delta t - V(t + \Delta t) & S(t+\Delta t) < S_{max} \\ S_{max} & \text{otherwise} \end{cases} . \qquad \textbf{(4.10)}$$

If the basin storage capacity is met fully within the interval (t, t+Δt) then V(t+Δt) is calculated from continuity as

$$V(t+\Delta t) = \pi_t \Delta t - (S_{max} - S(t)). \qquad \textbf{(4.11)}$$

A Pareto distribution was used to describe the distribution of soil depths for MODA The distribution function and probability density function for this distribution are:

$$F(c) = 1 - (1 - c/c_{max})^b \qquad 0 \le c \le c_{max}, \qquad \textbf{(4.12)}$$

$$f(c) = \frac{dF(c)}{dc} = \frac{b}{c_{max}}\left(1 - \frac{c}{c_{max}}\right)^{b-1} \qquad 0 \le c \le c_{max}. \qquad \textbf{(4.13)}$$

The parameter Cmax is the maximum storage capacity in the basin and b is the shape parameter controlling the degree of spatial variability of storage capacity over the basin.

These functions are illustrated in Figure 4.2 for a maximum storage capacity of 250mm. Not only is this the distribution that is most widely used in practice but, depending on the choice of shape parameter of storage capacity the distribution can be used to simulate a wide variety of catchment types. A very large value of b implies that the majority of stores are shallow stores whilst a very small value of b implies that the majority of store capacities are skewed towards Cmax. A uniform distribution of storage capacities is obtained as a special case when b=1. A constant storage capacity over the entire catchment equal to Cmax is obtained for b=0. For MODA the parameter b was left as a free parameter to be identified during calibration whereas a uniform distribution was assumed for MODB to reduce parameter identifiability problems within the model. The following relations can be derived for Pareto distributed storage capacities:

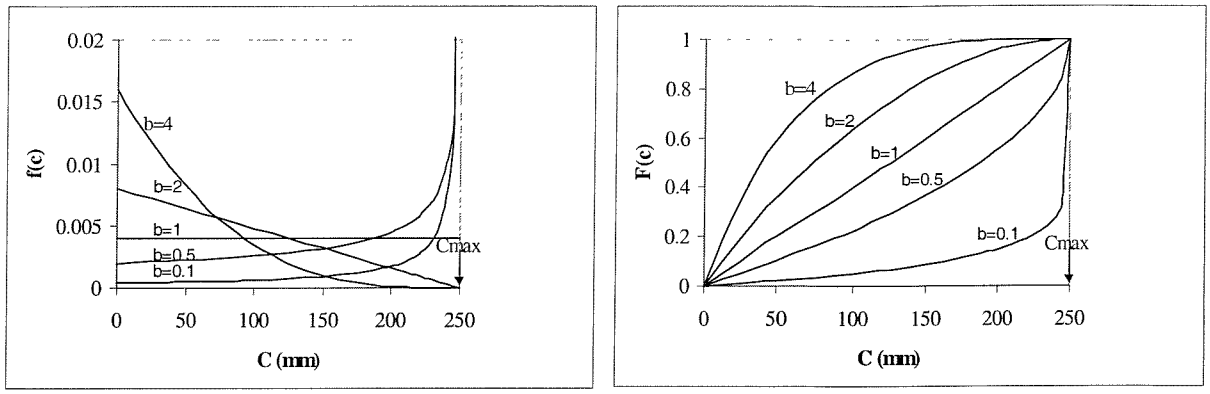
$$S_{max} = C_{max}/(b+1), \qquad \textbf{(4.14)}$$

$$S(t) = S_{max}\left\{1 - (1 - C^*(t)/C_{max})^{b+1}\right\}, \qquad \textbf{(4.15)}$$

$$C^*(t) = C_{max}\left\{1-(1-S(t)/S_{max})^{1/(b+1)}\right\}, \qquad \textbf{(4.16)}$$

$$v(t+\Delta t) = \pi\Delta t - S_{max}\left\{(1-C^*(t)/C_{max})^{b+1} - (1-C^*(t+\Delta t)/C_{max})^{b+1}\right\}. \qquad \textbf{(4.17)}$$

The relationship between rainfall and runoff implied by the above expressions, for given conditions of soil moisture, is presented in Figure 4.3




**(a) Probability density function** **(b) Distribution function**

**Figure 4.2 The Pareto distribution of storage capacity.**

Loss terms for the soil moisture store are the evaporation loss term ($E_i$) which is discussed in the subsequent section, the generation of direct runoff and, for MODA, a drainage loss term in which water from the PDSM is drained to the slow flow, or groundwater routing path. In this drainage term it is assumed that the rate of drainage over the interval, $d_i$, depends linearly on basin soil moisture content at the start of the interval where

$$d_i = \frac{S(t)}{k_g}. \qquad \textbf{(4.18)}$$

The constant of proportionality, 1/kg, is a drainage time constant with units of inverse time ($hr^{-1}$). With both losses to evaporation and recharge, the net rainfall, $\pi_i$, is defined as

$$\pi_i = P_i - E_i' - d_i. \qquad \textbf{(4.19)}$$

In this formulation, it arguable as to whether an interception module is necessary. The distinction that intercepted water is evaporated at the potential rate is only important when the SMD is limiting evaporation. Furthermore, this could probably be compensated for by the Cmax and b parameters. The interception model is potentially important when the drainage to the slow flow routing component is limited by soil moisture deficit. The incorporation of the interception store to intercept a fraction of the daily rainfall implies that a greater rainfall depth is required before the SMD is alleviated thus resulting in an increased drainage rate.

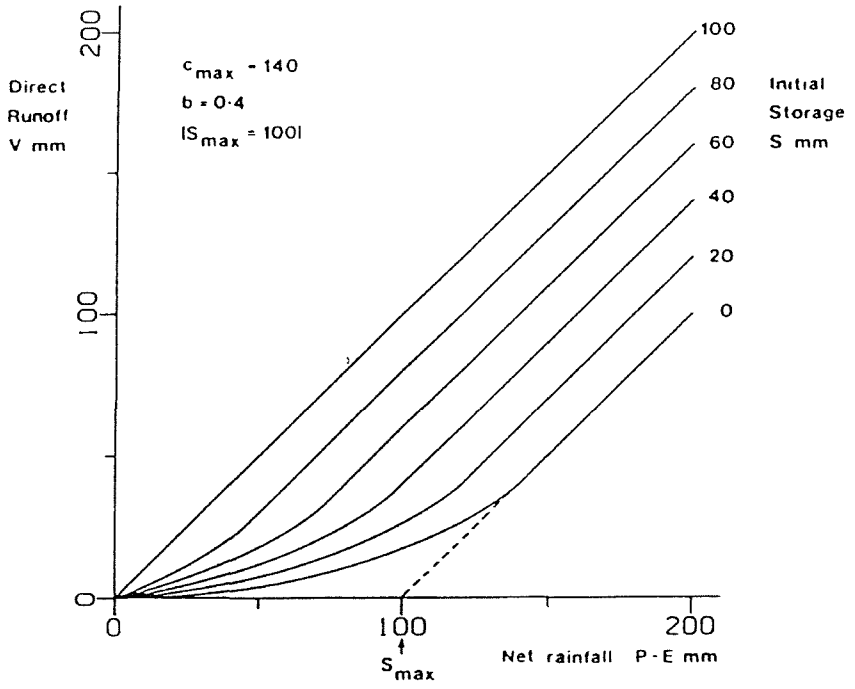



**Figure 4.3 Example relationships between rainfall and runoff for the Pareto distribution.**

In MODB it is assumed that there is no soil drainage, $d_i$. Direct runoff from the PDSM is split between a fraction $\beta$, which goes to groundwater storage and a fraction $(1-\beta)$ going to make up surface runoff.

## 4.2 THE EVAPORATION MODEL

Within the modelling framework developed for this study, the maximum evaporation that can take place in a time interval is equal to the Penman-Monteith based estimate of Potential Evaporation for that time step (Chapter 2).

Looking at evaporation from a process view-point it can be defined as the physical process by which water is converted from a liquid into a gas. The degree to which evaporation takes place is dependent on the vapour pressure of the air immediately above the evaporating surface. When the vapour pressure is low, water molecules can diffuse freely into the atmosphere. However, as the vapour pressure rises it becomes increasingly difficult for a net movement of water molecules into the atmosphere. At the saturation point it ceases entirely. Evaporation is therefore proportional to the vapour pressure deficit below the saturated vapour pressure at the air surface interface. The temperature of the atmosphere, the wind speed and vegetation canopy architecture, largely controls this deficit. The latter two factors control the rate at which water vapour is moved away from the air-surface interface and the degree of mixing which is related to turbulence over a canopy.

Evaporation from water, which is held on the catchment surfaces and water intercepted by vegetation evaporates at the potential rate defined by the atmosphere. Given an unlimited atmospheric demand, the amount lost from such sources is dependent on the amount held. For bare soils, the volume of surface storage is primarily dependent on small-scale topography and the infiltration capacity of the soil. For plants, canopy storage volume is dependent on leaf number, shape and size and canopy architecture.

The transfer of energy through the soil controls the rate at which evaporation can take place from within the soil. The soil matrix will also determine the rate of evaporation from the soil. The surface tension forces binding the water molecules to the soil particles (a function of size) and the tortuosity of the path connecting the water molecule to the soil surface primarily restrict this rate.

Evaporation from plant canopies is dependent on plant physiology. Evaporation takes place through leaf stomata and its magnitude is dependent on the size of the stomatal aperture. This is primarily dependent on climatic factors (temperature, vapour pressure, partial pressure of $CO_2$) and amount of water present in the plant. The amount of the water present in the plant is in turn dependent on the amount of the water in the soil and the ability of the vegetation to extract this water which is a function of the root structure and the type of soil. These complex processes controlling evaporation are represented within MODA by two conceptual processes:

- interception of precipitation and the subsequent evaporation at the potential rate determined by the atmosphere;
- evaporation from plant and catchment surfaces, which may take place at a rate equal to or less than the potential rate from the PDSM - with any reductions being a non-linear function of the soil moisture deficit within the PDSM.

As soil drainage is not considered in MODB the interception store was omitted from this model formulation. The modelling of interception losses and evaporation from the catchment surfaces is presented in the following sub-sections.

### 4.2.1 The Interception Model

The largest impact of changing the land use within a catchment, from a quantitative viewpoint, is the impact on evaporation processes. The largest modification of evaporation processes is generally associated with changes in interception losses, associated with vegetation architecture; leaf and stem structure. Enhancement of evaporation from these surfaces is primarily a function of aerodynamic resistance (Calder, 1990). The biggest increases in interception losses are associated with the transition from grass to coniferous afforestation. A number of approaches to the conceptual modelling of interception losses have been formulated (Rutter et al, 1971, Gash, 1979, Aston, 1979 and Calder 1986a & b). These models all use the same Penman Monteith equation to estimate evaporative losses from intercepted water but vary in the way the intercepted water is partitioned between different interception components. The interception model adapted for this study is based upon the daily interception model proposed by Calder (1986a). This model, which is the

simplest model, has been tested through observation on a number of vegetation types in the UK (Hall and Harding, 1993 and Harding *et al*, 1992). Considering vegetation class, j, covering a fraction of the catchment, $A_j$, the intercepted depth of rainfall on day i is given by

$$I_{ji} = A_j \gamma_j \{1 - e^{-\delta P_i}\}, \qquad \textbf{(4.20)}$$

where:

$I_{ji}$ = The interception depth within the day from vegetation class, j (mm);

$\gamma_j$= maximum daily interception loss for vegetation class, j (mm);

$\delta_j$= scaling constant for vegetation type, j ($mm^{-1}$);

$P_i$= precipitation depth within the day (mm).

Preliminary results, described in the subsequent section, demonstrate that interception losses can be adequately modelled (i.e. within experimental error) by re-parameterising this as a one-parameter model. This parameter, $\gamma$, " the maximum daily interception loss" is intrinsically related to vegetation type. The scaling constant, $\delta$, controls at what rainfall depth this maximum interception loss is reached, as demonstrated in Figure 4.4.

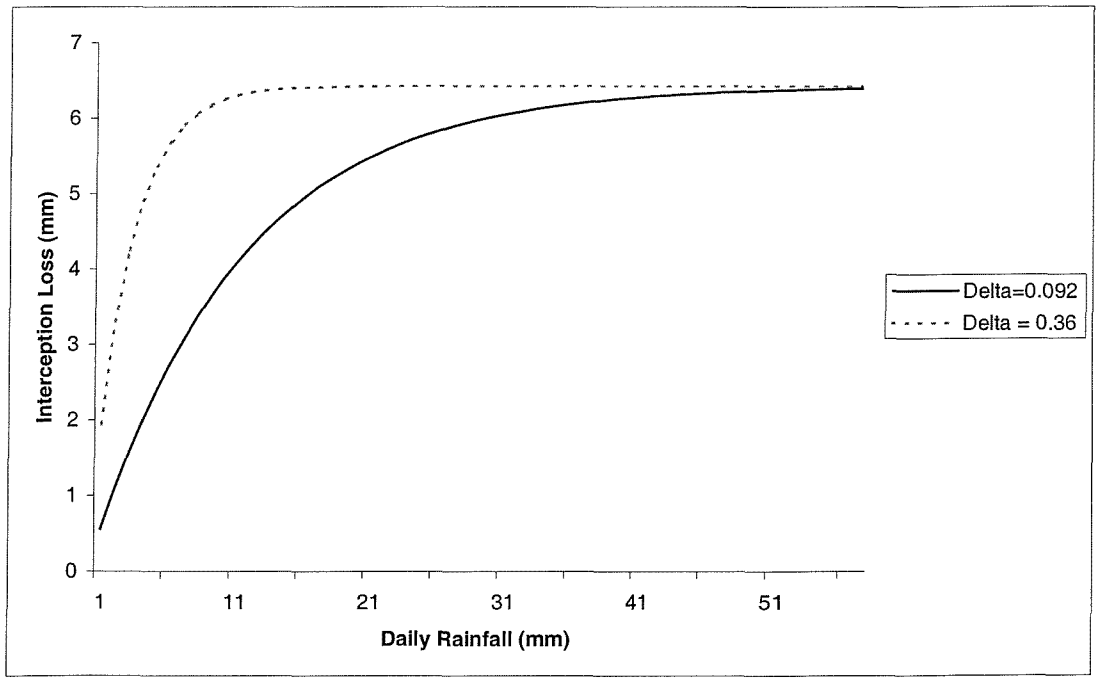



**Figure 4.4 The sensitivity of the interception model to the value of $\delta$ ($\gamma$ = 6.43 mm).**

The model has primarily been tested by previous researchers using data from interception losses associated with mature coniferous afforestation but has also been applied in the modelling of interception losses from heather and broadleaf, deciduous woodland. Published parameter values for the model are summarised in Table 4.1.

**Table 4.1 Published parameter values for the daily interception model**

| Source | Period | Interception Parameters | | |
|---|---|---|---|---|
| | | Fraction of annual Rainfall, $\alpha$ | $\gamma$ (mm) | $\delta$ ($mm^{-1}$) |
| *Coniferous Forest* | | | | |
| All sites: Dolydd Plynlimon, Crinan, Aviemore (Calder, 1986b) | | 0.35 | 6.99 | 0.099 |
| Plynlimon | 1974-1976 | 0.30 | 6.1 | 0.099 |
| Dolydd | 1981-1983 | 0.39 | 7.6 | 0.099 |
| Crinan | 1982-1984 | 0.36 | 6.6 | 0.099 |
| Aviemore | 1982-1984 | 0.45 | 7.1 | 0.099 |
| Balquidder-Kirkton (Calder et al, 1986a) | 1984-1985 | | 6.4 | 0.092 |
| *Heather* | | | | |
| Balquidder(Hall and Harding, 1993) | 1981 | - | 2.7 | 0.360 |
| Law's heather lysimeters (Calder et al, 1983) | | 0.16 | - | - |
| Sneaton moor lysimeter (Wallace et al, 1982) | 1980 | 0.19 | - | - |
| *Broadleaf Woodlands* | | | | |
| Beech annual (Harding et al 1992) | 1989-1991 | 0.14 | 2.1 | 0.099 |
| Ash annual (Harding et al, 1992) | 1989-1991 | | | |

As can be observed from this table the parameter values are relatively consistent for afforested catchments. Using daily rainfall data from 1961 to 1990 for Balquidder, it was found that the parameter values for the model, when applied within the Kirkton catchment (one of the pair of Balquidder catchments), can be recast as $\gamma$=6.2 mm and $\delta$=0.099 $mm^{-1}$. This does not introduce any bias in predicted daily losses and maintains an $R^2$ value of 1 between the two model configurations. The Balquidder calibrated model parameter values for heather are lower than the equivalent values for coniferous trees indicating that the

maximum interception loss from heather is lower but that a higher precipitation rate is required to achieve the maximum loss rate. Re-parameterising this equation with a value of $\delta$=0.099 $mm^{-1}$ yields a value of $\gamma$=4.1 mm without introducing any bias into predicted daily losses and maintaining an $R^2$ value of 0.91 between the two model configurations. It should be noted that the recast equation under predicts at low precipitation values, due to the lower value of $\delta$, and over predicts at high precipitation values due to the increased maximum daily interception loss.

The annual precipitation loss expressed as a fraction of the annual rainfall ($\alpha$) is 0.2 which is consistent with the published figures for heather listed in Table 4.1. It should be noted however, the average annual rainfall figures for the two published sites at 960mm $yr^{-1}$ for Stocks and approximately 1000 mm $yr^{-1}$ for the Sneaton Moor site are much lower than that of the Kirkton catchment.

Calder and Newson (1979), published a set of data presenting annual interception losses for a number of experimental sites afforested with coniferous trees with the annual interception losses presented as a fraction of annual rainfall. For this study M.Robinson (CEH-Wallingford) has made data from additional experimental sites available. These data are presented, together with the data of Calder and Newson in graphical form in Figure 4.5. The data are presented as annual interception loss data points, annotated with the name of the site, plotted as a function of average annual rainfall. The fraction of annual rainfall lost through interception increases as the annual rainfall decreases, which is as would be expected. Also presented in the figure are annual loss curves modelled using the daily interception model for:

- Coniferous Forest (calibrated)
- Heather (recast Balquidder)
- Heather (calibrated)
- Beech – annual (calibrated)
- Ash – annual (calibrated)

These loss curves were derived by applying the model to synthetic daily rainfall time series for the 10-year period 1980-89 covering the range of average annual rainfall from 485 mm/yr to 3000 mm/yr. These synthetic rainfall time series were constructed by taking a weighted averages of 1961-90 daily rainfall time series for the Box catchment in East Anglia (average annual rainfall of 495mm/yr) and the Balquidder catchments (average annual rainfall of 2400 mm/yr). Rainfall records from a wet catchment and a dry catchment were used to construct the synthetic rainfall time series (rather than just re-scaling a single rainfall time series) to try and capture the variations in rainfall intensity and number of rain days between wet and dry catchments. The percentage of rain days for the Box and Balquidder over the 10 year period were 59% and 71% respectively and the mean rainfall intensities over rain days were 2.2 and 9.2 mm/day.

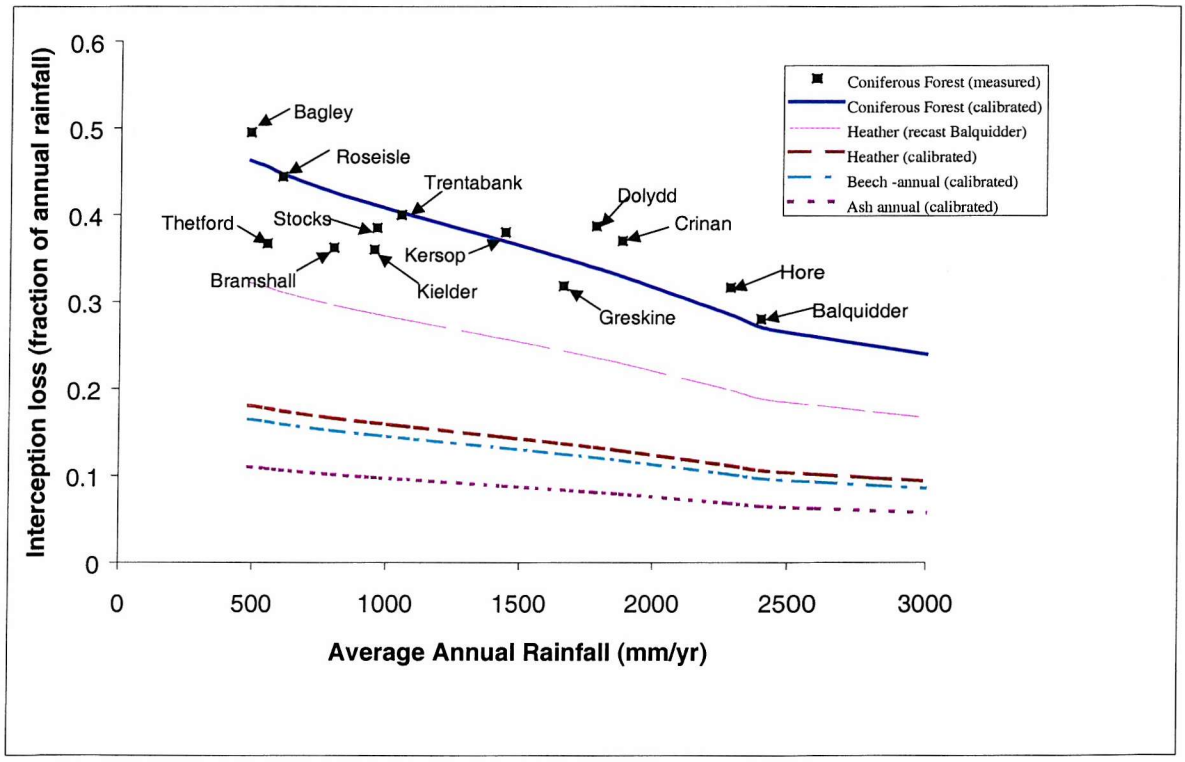



**Figure 4.5 Annual loss-rainfall curves for different vegetation types.**

Within the figure there is a breakpoint in the loss curves corresponding to an annual rainfall of 2400 mm/yr, above which the rainfall records were constructed solely from the Balquidder data. This indicates that the model is sensitive to the distribution of rainfall in terms of intensity and number of rain days. The maximum interception depth in the model will be exceeded more frequently for high intensity rainfall catchments with a greater

number of rain days. Above the 2400mm rainfall threshold, the number of rain days does not increase within the synthetic rainfall time series. The parameter values for the simulated loss curves are presented in Table 4.2 and were derived as follows:

- *Coniferous Forest (calibrated)*

This loss curve was derived by optimising the value of the maximum interception loss parameter by minimising the sum of standardised squared differences between the observed annual losses and those predicted by the model.

- *Heather*

Two loss curves were generated for Heather; one with the recast Balquidder parameters (Heather (recast Balquidder)) and one where the model was calibrated against the published annual interception loss values for the Stocks reservoir and Sneaton Moor lysimeter data (Heather Calibrated). The latter curve underestimates the experimental Balquidder data by 44%. As the average annual rainfall for Stocks and Sneaton Moor sites are more indicative of the average across the UK the calibrated Heather Curve is believed to be more representative.

- *Beech and Ash annual (calibrated)*

Beech and Ash are deciduous trees and consequently intercept more of incident precipitation when foliated during the summer months. This is offset by the fact that the seasonal distribution of rainfall means that winter precipitation is generally higher than that for summer. Harding et al (1992) developed seasonal model parameters for the daily interception model for Beech and Ash using experimental data from the Black Wood experimental site near Winchester, Hants over the period 1989-1991. The average annual rainfall over this period at Black Wood was 744mm/yr. For the purpose of this study, a single annual model is required for modelling interception losses. Annual models were developed by using a synthetic rainfall series with an annual mean of 744 mm/yr and optimising the model against the annual interception losses published by Harding *et al.* The weakness of these annual models is that they will underestimate the summer interception losses and over estimate the winter interception losses. It is interesting to note that the loss curves lie slightly below those for heather.

**Table 4.2 Parameter values for simulated loss curves**

| Vegetation type | Parameters | |
|---|---|---|
| | $\gamma$ (mm) | $\delta$ ($mm^{-1}$) |
| Coniferous Forest | 5.9 | 0.099 |
| Heather (recast Balquidder) | 4.1 | 0.099 |
| Heather (calibrated) | 2.6 | 0.099 |
| Beech – annual (calibrated) | 2.1 | 0.099 |
| Ash-annual (calibrated) | 1.4 | 0.099 |

One criticism of the interception model is that it assumes that all the intercepted water within a time step is evaporated and thus it does not necessarily maintain continuity between time steps. To address this the model was modified for use within MODA by the inclusion of an interception store equal to the maximum interception depth, $\gamma$. The conceptual structure of the interception store is presented in Figure 4.6.

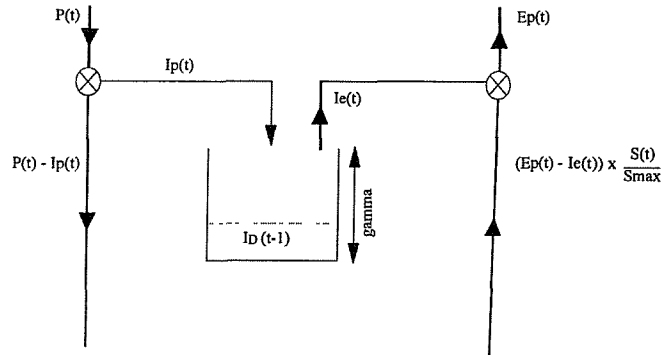



**Figure 4.6 The conceptual structure of the interception module.**

Taking the potential depth of rainfall in time step $(t+\Delta t)$ that could be intercepted from the precipitation depth within the time step, $Ip_{(t+\Delta t)}$, the potential depth of water stored in the interception store $I_{(t+\Delta t)}$ is calculated from continuity as

$$I'_{(t+\Delta t)} = I_t + Ip_{(t+\Delta t)} - PE_{(t+\Delta t)} \cdot \qquad (4.21)$$

For the condition $I'_{(t+\Delta t)} \leq 0$, the actual depth of water intercepted, $I_{(t+\Delta t)}$, is zero, and the residual potential evaporation demand to be met from the soil moisture store, $PE'_{(t+\Delta t)}$, and the residual precipitation input into the soil moisture store, $P'_{(t+\Delta t)}$, are given by:

$$PE'_{(t+\Delta t)} = PE_{(t+\Delta t)} - (I_t + Ip_{(t+\Delta t)}),$$

$$P'_{(t+\Delta t)} = P_{(t+\Delta t)} - Ip_{(t+\Delta t)}.$$

If the $I'_{(t+\Delta t)}$ lies in the interval $[0,\gamma]$ then:

$$PE'_{(t+\Delta t)}=0,$$

$$I_{(t+\Delta t)} = I'_{(t+\Delta t)},$$

$$P'_{(t+\Delta t)} = P_{(t+\Delta t)} - Ip_{(t+\Delta t)}.$$

If $I'_{(t+\Delta t)}$ is greater than $\gamma$ then $PE'$ is equal to zero, $I_{(t+\Delta t)}$ is equal to $\gamma$ and

$$P'_{(t+\Delta t)} = P_{(t+\Delta t)} - Ip_{(t+\Delta t)} + I'_{(t+\Delta t)} - \gamma.$$

### 4.2.2 Estimation of Evaporative Losses from the PDSM

If the potential evaporative demand within a time step is not met by evaporating water from the interception store, the residual evaporative demand is met from the PDSM. The dependence of the evaporation rate on soil moisture content within the PDSM is introduced by assuming a simple Pareto distribution function between the ratio of actual to residual potential evaporation, $E_i/PEr_i$, and soil moisture deficit, Smax - S(t):

$$\frac{E_t}{PEr_t} = 1 - \left\{ \frac{S_{\max} - S(t)}{S_{\max}} \right\}^{be}. \qquad \textbf{(4.22)}$$

A plot of $E_i/PEr_i$ as a function of soil moisture deficits is presented in Figure 4.7 for a range of values of the shape parameter, be. For a value of be=6 this figure demonstrates that evaporation takes place at the potential rate until an SMD of 50% is reached. Beyond this point the evaporation rate reduces below the potential rate. The rate of change of this reduction rapidly increases with increasing SMD.

In the formulation of the PDSM Smax is equal to the mean store depth, $\bar{c}$. Consider the special case of the Pareto distribution, b=0, in which the PDSM is represented by a single depth, as defined by Cmax, in this case Smax =Cmax. In this case the formulation of the evaporation model is similar to the Penman Drying Curve model (Penman, 1949, Grindley, 1970) implemented within the Thames Catchment Model (Chapter 3). The parameter Cmax, under these circumstances, can be thought of as being equivalent to twice the "rooting depth" within the Penman model. The rooting depth is an effective depth, which is a function of the root structure of vegetation and the soil. Vegetation can easily extract water from the soil up to an SMD equivalent to the rooting depth, water is not limiting and evaporation takes place at the potential rate. As the SMD increases beyond the rooting depth the extraction of water becomes harder and thus the evaporation rate decreases accordingly.

In the original Penman model a constant reduction factor is used once the SMD exceeds the rooting depth and the vegetation is can only extract a further 25mm of water until extraction becomes minimal and evaporation ceases. The model used in this study allows evaporation to continue, albeit at an ever decreasing rate, until the SMD approaches twice the rooting depth.

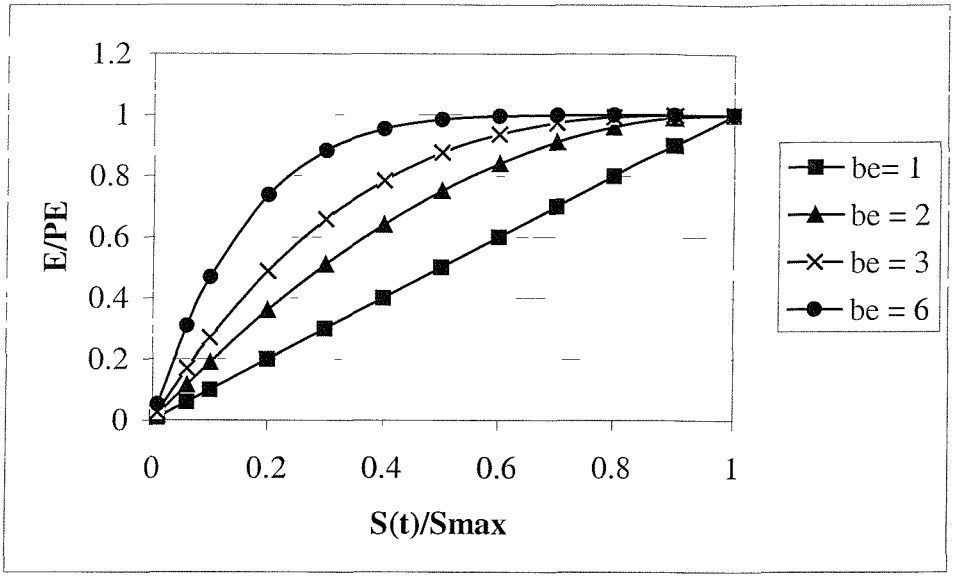



**Figure 4.7 The sensitivity of the evaporation model to the value of $b_e$.**

## 4.3 ROUTING MODEL

As discussed in the previous sections the PDSM partitions effective rainfall into direct runoff, groundwater recharge and soil moisture storage. Direct runoff is routed through the surface storage, or quick response path. The effective rainfall partitioned into groundwater recharge is routed through a subsurface storage or slow response path. Both routing systems can be defined by a variety of Hortonian non-linear storage reservoirs (Horton,1938). The basic form for a non-linear reservoir defining the outflow at a point in time, q(t), is given by

$$q(t) = \frac{1}{k} s(t)^n, \qquad \textbf{(4.23)}$$

where:

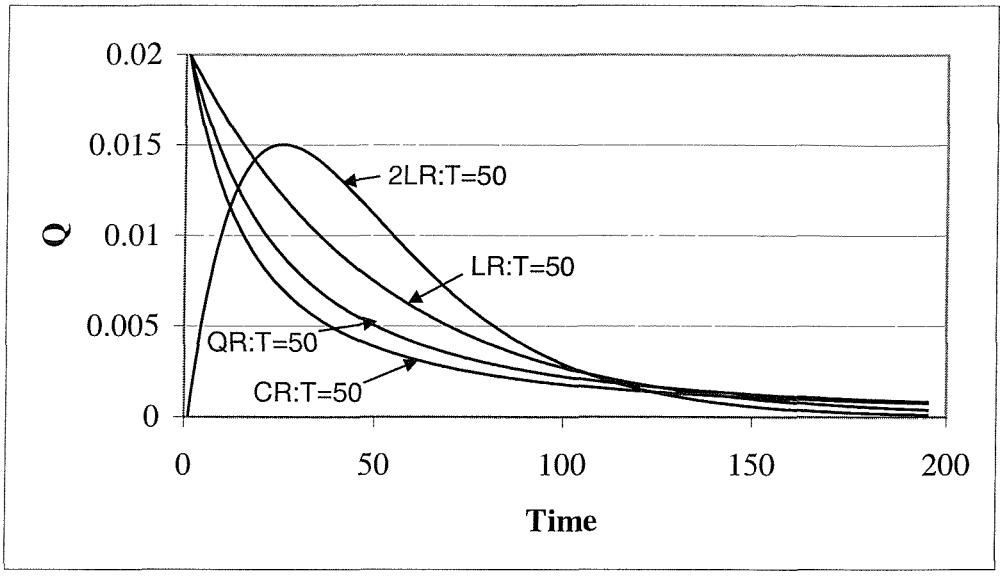
s(t) = the volume of water in storage at time,t;

k = a constant (with units of time);

n = the order of the reservoir.

The response of a linear (n=1), quadratic (n=2) and cubic (n=3) non-linear reservoirs, with the same time constant (50 time units) to a unit impulse is presented within Figure 4.8. The response of a cascade of two linear reservoirs in series is also presented. As can be observed from this figure the non-linear reservoirs have a higher initial recession rate than

the linear reservoir but in the tail of the recessions the rates are fairly similar. Horton noted the insensitivity of response to the value of the exponent n, provided k could be adjusted to compensate. In rainfall runoff modelling it is common practice to choose an appropriate value of n, and to optimise k to avoid the problem of covariance between k and n . As the focus of the study is on modelling the lower parts of recession curve accurately the choice of configuration is not critical. A linear reservoir was selected to represent the routing for both the quick and slow routing paths. Selection of the same configuration for both paths facilitates a direct comparison of time constant when applying the model to a catchment.




R= reservoir, L=linear, Q=quadratic, C=cubic, 2L = two linear reservoirs in series

**Figure 4.8 Response curves for routing reservoirs.**

The explicit formulation of equation 4.24 neglects that, within a time step, the instantaneous value of s is dependent on the function of the outflow, q. Combining 4.24 with the continuity equation:

$$\frac{dS}{dt} = u - q, \qquad \textbf{(4.24)}$$

in which u is the inflow over the time period yields for a linear reservoir:

$$\frac{dq}{dt} = \frac{1}{k}(u - q) \; . \qquad \textbf{(4.25)}$$

Rearranging 4.26 and integrating over the time period (t,t+Δt) gives the explicit recursive solution for q as

$$Q_{(t+\Delta t)} = e^{-\frac{\Delta t}{k}} q_t + u\left(1 - e^{-\frac{\Delta t}{k}}\right). \qquad \textbf{(4.26)}$$

The parameters for MODA and MODB are summarised in Table 4.3. This table also lists units, where applicable. For convenience the units for K1 and Kb were set as hours.

**Table 4.3 Model Parameters**

| Parameter name | Unit | Description |
|---|---|---|
| *Interception store* | | |
| γ | mm | Depth of interception store in MODA |
| *Probability-distributed soil Moisture store* | | |
| Cmax | mm | The maximum store capacity within the catchment. |
| b | none | The exponent of the Pareto distribution, controlling the spatial variability of store capacities. |
| kg | hour | The groundwater recharge time constant for MODA. |
| β | none | The runoff factor, which controls the split of direct runoff between surface and groundwater storage routing for MODB. |
| *Routing Module* | | |
| K1 | hour | The time constant for the quick flow linear reservoir. |
| Kb | hour | The time constant for the slow flow, or base flow linear reservoir |

# 5 The selection of catchment data sets and derivation of catchment characteristics

The development of relationships between model parameters and/or parameter vectors and catchment characteristics has to be based on catchments with good quality data and relatively natural flow regimes as the river flows must approximate a natural response to the physical properties of the catchment. This chapter discusses the selection of usable gauged catchments and the development of a catchment characteristics database describing the climate, soils, land use and topographical variations within the useable catchments.

## 5.1 DERIVATION OF USABLE CATCHMENT DATA SET

### 5.1.1 Initial catchment classification

The selection of catchments built upon the work of Gustard *et al* (1992). In this study all of the stations on the National River Flow Archive (held at CEH-Wallingford) were graded using a hydrometric classification and a degree of artificial influences classification based on the data available up to the end of 1989.

The hydrometric classification (A, B or C) was based upon a Low flows Sensitivity Index (SI) that measured the percentage change in the Q95 flow represented by a +10mm stage variation above the Q95 stage and a Factorial standard error of estimate at the Q95 stage. This error statistic describes the scatter of check gaugings about the low flow discharge, derived from the theoretical or empirical rating. This was derived from actual data or more commonly derived from graphs of rating curves with superimposed check gaugings. At purpose-built structures with no check gaugings available, a theoretical estimate based on the probable error in deriving a gauged flow from the head was used. In addition to these quantitative measures, qualitative information regarding station maintenance, gauge bypass, etc. was also used.

The artificial influence classification (A, B, C) was based upon an assessment of the impact of artificial influences on the ratio of the natural Q95 and the natural mean flow (MF). Essentially the assessment procedure identified those catchments with a biased value of Q95

due to a nett loss (abstraction) or gain (effluent returns) to the catchment. This was calculated by semi naturalising the flow statistics, using data collated on major artificial influences within the catchments. The influence considered were abstractions (with licenced maximum abstraction rates greater than 10 $ls^{-1}$), consented discharges (for all discharges greater than 10 $ls^{-1}$), reservoir yield and compensation flows for reservoirs exceeding capacity of 500 ML. Information on abstraction licences and discharge consents were provided by the regions of the Environment Agency. Data on reservoir yields and compensation flows were extracted from Gustard *et al.* (1987). The major limitations of this method with respect to the current study are that the approach sought to maintain the ratio of Q95/MF, information about artificial influences were authorised quantities (rather than actual) and in practice not all artificial influences will have been included.

**Table 5.1 Classification scheme for low flow suitability grading *(source: Gustard et al, 1992)***

**Classification of hydrometric quality**

GRADE A
Accurate low flow measurement over a sensitive control (Sensitivity index, SI less than 20%) with the scatter of spot gaugings about the rating curve at the Q95 discharge having a factorial standard error of estimate of less than 1.1, and no obvious deterioration of the gauging station due to siltation, weed growth or vandalism

GRADE B
Less accurate low flow measurement with either a less sensitive control (SI between 20% and 50%) or a factorial standard error of estimate of between 1.1 and 1.2, and/or observed periodic deterioration of the gauging station due to siltation, weed growth or vandalism.

GRADE C
Station with low accuracy of low flow measurement due to either an insensitive control (SI in excess of 50%), and/or with the scatter of gaugings about the rating curve at the Q95 discharge having a factorial standard error of estimate in excess of 1.2, and/or observation of sustained deterioration of the gauging station due to siltation, weed growth or vandalism.

GRADE U Unclassifiable due to insufficient information.

**Classification of degree of artificial influence**

GRADE A
The gauged Q95/mean flow ratio differs by less than 20% from the estimated natural Q95(1)/mean flow ratio.

GRADE B
The gauged Q95/mean flow ratio differs by more than 20% but less than 50% from the estimated Q95(1)/mean flow ratio.

GRADE C
The gauged Q95/mean flow ratio differs by more than 50% from the estimated Q95/mean flow ratio.

GRADE U
Unclassifiable due to insufficient information.

The criteria used for classifying gauging stations according to hydrometric quality at low flows and the degree of artificial influences are summarised in Table 5.1. For this study the pool of gauged catchments was expanded to take into account both new gaugings stations and those catchments whose records were too short (<6 years) to be considered by Gustard *et al* (1992). An alternative assessment of these stations was undertaken. Hydrometric quality was assessed using the following criteria.

- The Environment Agency's internal procedure for hydrometrically classifying gauging stations. The Agency Regions provided a list of hydrometric grades derived using their internal classification method.
- Use of the Sensitivity Index. Some station files, held on the National Water Archive at CEH Wallingford, have a rating equation from which the sensitivity index was calculated.
- Qualitative Gauging station descriptions. Gauging station descriptions are held on the National Water Archive. Whilst the descriptions do not providing a means by which the station may be graded, they provide useful information of the station in terms of rating, accuracy of the station and factors which may affect that accuracy, such as weed growth. This information may be used to qualitatively determine the usability of a station.

The artificial influence grading was based upon the catchment descriptions, held on the National Water Archive, describe the nature of the artificial influence in the catchment. This method of identifying usable gauging stations is useful because catchments without any influence are described as 'natural'. However, it is more difficult to distinguish between those stations that may be either B or C graded. Lists of these additional stations (with provisional grades) were circulated to Environment Agency and SEPA Regions for comment. Feedback was received from the North West region of the Environment Agency and the North and East regions of SEPA. This feedback was used to revise grades accordingly.

### 5.1.2 Catchment selection

Catchments with significant influences, and/or significant hydrometric errors at low flows were excluded from this study. This was achieved by initially selecting those stations graded AA under the classification scheme discussed above. Stations from Northern Ireland were omitted because of the lack of MORECS PE data for Northern Ireland (Chapter 2). For the additional stations it was assumed that these were of AA quality unless there was evidence to the contrary. This initial selection reduced the number of candidate stations to 437. A major weakness of the artificial influence classification scheme, from the perspective of this study, was that the classification only considered the impact on the Q95/MF ratio. This was addressed by extracting the estimated artificial influence quantities from the original paper records of Gustard *et al* (1992). The selected stations were then re-assessed and stations where the nett influence was greater than 10% of the naturalised MF or 20% of the naturalised Q95 flow were excluded. Stations where the contiguous period of record was less than 12 years were also excluded. This latter criterion was used because a minimum calibration record length of 10-years (with a 2 year model 'warm up') was to be used for the modelling studies. A missing data criterion was applied to reject years if more than 30 days of data were missing. The use of this criterion reduced the number of candidate gauging stations for the modelling work to 318. The location of these stations is presented in Figure 5.1.

There are greater numbers of candidate catchments in the wetter west and north of the UK (Figure 5.1). This is confirmed by the mean and 95% limit (95% of catchments have values between these limits) summary statistics for gauged Average Annual Runoff (AARO) and BFI derived from the gauged flow presented in Table 5.2. The dimensionless Base Flow Index is a useful measure of how permeable a catchment is; a value near unity is indicative of large, groundwater derived base flow, a small value indicates a flashy, impermeable catchment. The derivation of BFI is discussed in more detail within chapter 6. The values of BFI in Table 5.2 indicate that the range of catchment types seen in the UK is represented by the data set.

**Table 5.2 Summary statistics for candidate catchments**

| Statistic | AARO (mm/yr.) | BFI |
|---|---|---|
| 95% u.l. | 2300 | 0.92 |
| Mean | 650 | 0.46 |
| 95% l.l. | 117 | 0.19 |

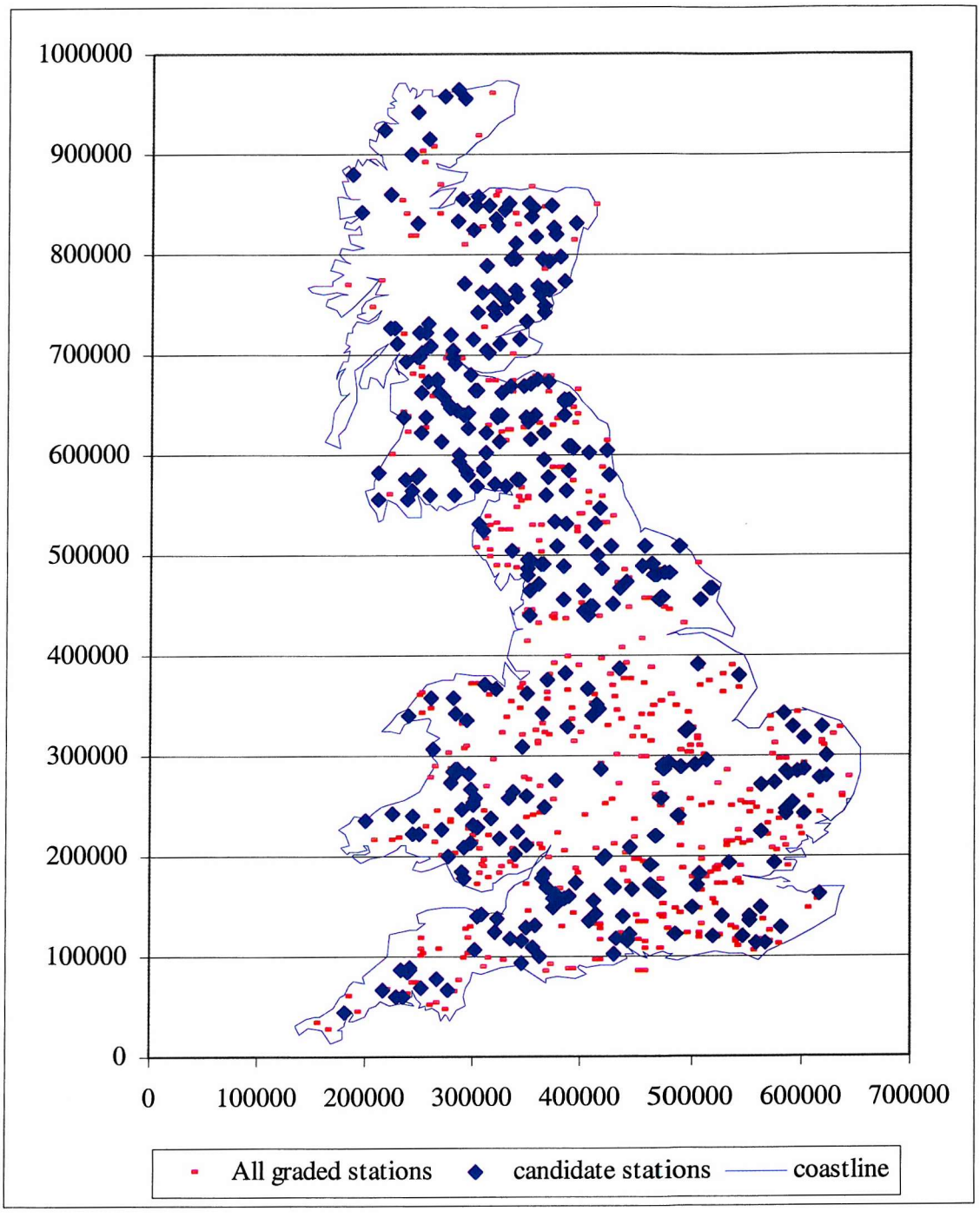



**Figure 5.1 The locations of the gauging stations for the candidate catchments.**

## 5.2 THE CATCHMENT CHARACTERISTICS

The catchment characteristics used in this study can be grouped into three classes:

- Topographical characteristics
- Climatological characteristics
- Soils and Land cover

These are described in the subsequent sub-sections.

### 5.2.1 Topographical characteristics

The topographical catchment characteristics were derived using the former Institute of Hydrology's Digital Terrain Model (DTM) (Morris and Flavin, 1990). This DTM consists of five 50m-resolution grids, the three that have been used to derive topographical characteristics are:

- Altitude grid – the altitude of each node above mean sea level, derived from Ordnance Survey contour data using an interpolation procedure described by Morris and Flavin, (1990).

- Drainage direction grid – from the altitude data the gradient between a node and its nearest eight neighbours is calculated and the drainage direction is taken as the steepest "down slope" gradient to it's nearest neighbour.

- Inflow grid – This grid identifies for each node, as an eight-bit code, the number of the eight nearest neighbour nodes that drains towards it, effectively defining a 50 m grid of drainage areas.

Catchment area estimation

The accurate definition of the catchment boundary and hence the area draining to each catchment is extremely important. The area is an *a priori* model parameter used in the modelling studies and the boundary is used to identify the extent of each catchment characteristic grid within a catchment boundary. The DTM was used to generate catchment

boundaries and areas for each of the study catchments using the method described by (Morris and Heerdegen, 1988). A comparison was then made between the estimates of area derived using the DTM and the manually derived catchment area estimates held for most gauged catchments on the National River Flow Archive.

The same exercise has been undertaken for the catchment data set used in the development of the Flood Estimation Handbook (FEH) (Bayliss, 1999). The results of this analysis were used for catchments that occur within the data set used for this study and extended to cover those catchments not used within the FEH. Bayliss identified that only 5% of FEH catchments differed in area by 10% or more. Some of these catchments have boundaries that, through drainage diversion, do not always follow the topography, which the DTM-derived watershed must always do. In other cases, the generation of DTM flow paths has been flawed by difficulties encountered when using digitised rivers to fix the location of valleys (a key element of the generation of the DTM). Where initial estimates differed by more than 10% it was possible, for a number of catchments, to correct the DTM where it had chosen an incorrect stretch. Where the differences in catchment area were greater than 10% and could not be resolved the catchments were excluded from the initial data set of AA graded stations. Therefore, for all of the catchments used in the study the error between DTM generated catchment areas and manually derived catchment areas is less than 10%.

Flood Estimation handbook catchment descriptors

A set of topographical catchment characteristics has been derived for the FEH. These are described fully by Bayliss (1999). Using FEH nomenclature these are called catchment descriptors. To differentiate between these descriptors and characteristics for this study the term descriptor is defined as a catchment property that is function of the structure of a specific catchment and cannot be generalised as a grid. The catchment descriptors that were thought to be potentially important in controlling the variability in daily flows and hence the parameters of a calibrated rainfall runoff model were used for this study. These descriptors are summarised in Table 5.3.

ASPVAR is derived using circular statistics (Mardia, 1972). Each slope direction within a catchment is resolved as x and y components. ASPVAR is the resultant of the mean x and y components.

**Table 5.3 Glossary of FEH catchment descriptors**

| Catchment descriptor | Units | Description |
|---|---|---|
| ALTBAR | m | The mean altitude of the catchment |
| ASPBAR | Degrees (0,360=Nth) | The mean direction of all 50m slopes in the catchment. Represents the dominant aspect of catchment slopes |
| ASPVAR | none | The invariability of slope direction. Values approaching one indicate dominance of one direction |
| DPLBAR | km | The mean of the distances measured between each node (on regular 50-m grid) and the catchment outlet. Characterizes catchment size and configuration. |
| DPLCV | km | The CV of the distances measured between each node and the catchment outlet. Descriptor of drainage path configuration |
| DPSBAR | m/km | The mean of all the inter-nodal slopes for the catchment. Characterizes the overall steepness within the catchment |
| LDP | km | The longest drainage path defined by measuring the distance from each node to the defined catchment outlet. Principally a measure of catchment size but also reflects catchment configuration |

### 5.2.2 Climatological characteristics

Standard period 1961-90 Average Annual and monthly Rainfall

The generation of the Meteorological Office 1961-90 SAAR 1km grid has been described in Chapter 2. Monthly 61-90 1km-rainfall grids were derived from the Meteorological Office daily rainfall archive using the modified triangular planes methodology described in Chapter 2. The average annual and monthly rainfall statistics were derived for all catchments using the catchment boundaries described in section 5.2.1. Monthly rainfall statistics were used to derive the PP catchment characteristic.

Standard period 1961-90 Average Annual and monthly Potential Evaporation

The development of a 1km-resolution grid of Penman Monteith estimates for short grass is described in Chapter 2. The same methodology was used to generate standard period monthly PE estimates. These estimates were used in conjunction with the monthly rainfall statistics to derive the PP catchment characteristic.

Potential for Soil Moisture Deficits (PP)

Soil moisture deficits potentially occur in a catchment when the evaporative demand and drainage from the soil exceeds the incident precipitation. As significant soil moisture deficits build up the rate at which water evaporates reduces. This process is described in more detail within Chapter 4. The PP characteristic was developed to represent this process in a relatively crude way. The difference between the catchment average monthly rainfall and potential evaporation was calculated for each month within the year. The difference was summed for months in which it was negative (potential evaporation demand exceeds precipitation) and express as a fraction of the annual potential evaporation estimate. The interpretation of the PP statistic is that is represents that fraction of the potential evaporation demand that might occur when water availability is limited. The larger the value of PP for a catchment, the more likely there are to be significant soil moisture deficits occurring within the catchment.

### 5.2.3 Land cover and soils

Hydrology of Soil Types Classification, BFIHOST, SPRHOST and HOSTRES

The Hydrology of Soil Types (HOST) project is a soil association (mapped at a scale of 1:250,000) based hydrological response classification (Boorman *et al*, 1995) of soils across the United Kingdom.

The HOST classification was developed by grouping soil associations into self-similar groups based upon their physical properties. The physical properties considered are presented in Table 5.4. The association of HOST classes with hydrogeological units is presented in Table 5.5. The depths to an aquifer or groundwater are based on observed and estimated depths for each soil series. Thus although the HOST project makes extensive use of a soil database it does incorporate hydrogeological data.

**Table 5.4 Physical properties of soil series used in HOST classification *(source: Boorman et al, 1995)***

*1 Soil hydrogeology*
A soil hydrogeology classification was derived specially for the HOST project using soil parent material definitions and the 1 625,000 scale Hydrogeological Map of England and Wales (Institute of Geological Sciences 1977) The scheme is used to differentiate between mechanisms of vertical water movement (e g intergranular or fissure flow), and to distinguish between permeable, slowly permeable and impermeable substrates Definitions of permeability are based on Bell (1985) Permeable substrates have a vertical saturated conductivity of more than 10cm/day and an aquifer or shallow water table Slowly permeable substrates have a vertical saturated conductivity of 10 to 0 1cm/day and may contain a local or concealed aquifer Impermeable substrates have a vertical saturated conductivity of less than 0 1cm/day and contain no aquifers

*2 Depth to aquifer or groundwater*
This indicates the time taken for excess water to reach the water table

*3 Presence of a peaty topsoil*
A raw peaty subsoil indicates saturated surface conditions for most of the year, limits infiltration and provides lateral pathways for rapid response in the uppermost part of the soil
*4 Depth to a slowly permeable layer*
A slowly permeable layer impedes downward percolation of excess soil water causing periodic saturation in the overlying layer Storage is reduced and there is an increased response to heavy rainfall
*5 Depth to gleyed layer*
Gleying is the presence of grey and ocherous mottles within the soil caused by intermittent waterlogging The particular definition of gleying used (Hollis 1989) identifies soil water layers wet for at least 30 days each year, or soils that are artificially drained
*6 Integrated air capacity*
The air capacity, or 'drainable' pore space of a soil layer, is defined as its volumetric air content at a tension of 5 Kilo Pascals (KPa) (approximately field capacity) Integrated air capacity (IAC) is the average percentage air volume over a depth of one metre This provides a surrogate for permeability in permeable soils and substrates (Hollis and Wood 1989) In slowly permeable soils or impermeable soils and substrates, IAC indicates the capacity of a soil to store excess water

Simple conceptual models describing the flow paths of water provided a structure to the classification scheme. Initially the 969 soil series were analysed and those with similar flow paths (indicated by their physical properties) were grouped together into a single HOST class. This produced a more manageable data set for further analysis. The percentage cover of the reduced number of classes were then related to gauged BFI values using multiple regression analysis, and by inspection of the response of individual catchments. The regression analysis provided further guidance on discriminating and grouping soil series. The process resulted in a final 29-class system. The classification is summarised in Table 5.6 in which classes grouped by physical characteristics.

Table 5.7 presents the fractional extent of the 29 HOST classes within the United Kingdom. The table also presents the percentage of the candidate catchments containing each HOST class and the mean fractional extent within those catchments. The spatial extent of urban coverage and lakes is not included as the catchments used within the study had a negligible urban component and did not contain major surface water bodies. The

table highlights that not only are some HOST classes not well represented within the catchment data set but also that these classes are not well represented in the UK. Another issue is that a small amount of some HOST classes occur in a large number of catchments. However, subject to these provisos, the HOST classes observed within the UK are well represented within the sample of catchments selected for use within this study.

**Table 5.5 Hydrogeological units within HOST classes** ***(source: Gustard et al, 1992)***

| Hydrogeological unit | HOST classes |
|---|---|
| 1 Soft sandstone, weakly consolidated sand | 2, 9, 12, 14 |
| 2 Weathered/fissured intrusive/metamorphic rock | 3, 8, 11, 12, 13, 14 |
| 3 Chalk, chalk rubble | 1. 8. 12 |
| 4 Soft Magnesian, brashy or Oolitic limestone and ironstone | 12, 29 |
| 5 Hard fissured limestone | 3, 14 |
| 6 Hard coherent rocks | 18, 21, 26 |
| 7 Hard but deeply shattered rocks | 7, 8, 14, 16 |
| 8 Soft shales with subordinate mudstones and siltstones | 17, 20, 23, 25 |
| 9 Very soft reddish blocky mudstones (marls) | 17, 20, 23 |
| 10 Very soft massive clays | 19, 22, 24 |
| 11 Very soft bedded loams, clays and sands | 15, 17, 23 |
| 12 Very soft bedded loam/clay/sands with subordinate sandstone | 15, 17, 23, 25 |
| 13 Hard fissured sandstones | 3, 12, 14 |
| 14 Earthy peat | 10 |
| 15 River alluvium | 7, 8, 9, 11 |
| 16 Marine alluvium | 7, 8, 9 |
| 17 Lake marl or tufa | 9 |
| 18 Colluvium | 5, 13, 14, 16 |
| 19 Blown sand | 4, 6, 9, 11 |
| 20 Coverloam | 5, 7, 8, 9, 12, 13 |
| 21 Glaciolacustrine clays and silts | 17, 23, 25 |
| 22 Till, compact head | 8, 15, 17, 20, 23, 25 |
| 23 Clay with flints or plateau drift | 1, 17, 23, 25 |
| 24 Gravel | 4, 5, 8, 9, 11, 14 |
| 25 Loamy drift | 5, 7, 8, 9, 12, 13, 14, 16 |
| 26 Chalky drift | 1, 7, 9, 12 |
| 27 Disturbed ground | 20, 23 |
| 34 Sand | 4, 6, 9 |
| 35 Cryogenic | 16 |
| 36 Scree | 8 |
| 43 Eroded blanket peat | 27 |
| 44 Raw peat | 11, 28 |

**Table 5.6 The HOST classification *(Source: Boorman et al 1995)***

| SUBSTRATE HYDROGEOLOGY | MINERAL SOILS | | | | | | PEAT SOILS | |
|---|---|---|---|---|---|---|---|---|
| | Groundwater or aquifer | No impermeable or gleyed layer within 100 cm | Impermeable layer within 100 cm | OR gleyed layer within 40 cm | Gleyed layer within 40 cm | | | |
| Weakly consolidated, microporous bypass flow uncommon (Chalk) | Normally present and at > 2m | 1 | 12 | | 13 | | 14 | |
| Weakly consolidated microporous, bypass flow uncommon (Limestone) | | 29 | | | | | | |
| Weakly consolidated, macroporous, by pass flow uncommon | | 2 | | | | | | |
| Strongly consolidated, non or slightly porous Bypass flow common | | 3 | | | | | | |
| Unconsolidated macroporous bypass flow very uncommon | | 4 | | | | | | |
| Unconsolidated microporous, bypass flow common | | 5 | | | | | | |
| Unconsolidated, macroporous bypass flow very uncommon | Normally present and at < 2m | 6 | | | IAC < 12 5 [< 1m day$^{-1}$ ] | IAC > 12 5 [> 1m day$^{-1}$ ] | *Drained* | *Undrained* |
| Unconsolidated, microporous, bypass flow common | | 7 | | | 8 | 9 | 10 | 11 |
| Slowly permeable | No significant groundwater or aquifer | 15 | IAC > 7 5 17 | IAC < 7 5 20 | 23 | | 25 | |
| Impermeable (hard) | | 16 | 18 | 21 | | | 26 | |
| Impermeable (soft) | | | 19 | 22 | 24 | | | |
| Eroded Peat | | | | | | | 27 | |
| Raw Peat | | | | | | | 28 | |

Deriving multivariate statistical relationships between a dependent variable and the fractional extent of HOST classes can give rise to parameter identifiability problems for poorly represented classes. In developing the BFI model (used to aid classification) Boorman *et al* (1995) resolved this problem by using a bounded regression approach. An unbounded regression yielded parameter estimates in excess of unity and parameter estimates of high uncertainty. The bounds placed on the maximum and minimum parameter estimates for a HOST class was based on the conceptual models underlying the HOST classification, and effectively maintained realistic parameter estimates. A regression equation derived from the catchment data set would have difficulty estimating the BFI parameters associated with poorly represented HOST classes such as HOST11 and HOST20. A bounded regression approach effectively applies conceptual knowledge of the relative magnitudes of the model parameters prior to the derivation of model coefficients.

The BFI model developed by Boorman *et al* was used to generate the BFIHOST catchment characteristic for this study. Boorman *et al* also developed a model for estimating the Standard Percentage Runoff (a measure of the percentage of rainfall that generates runoff (NERC, 1970)). This second model was used to verify the HOST classification. A third

characteristic HOSTRES was also derived based on HOST. This characteristic is the residual from a linear regression model fitted between SPRHOST and BFIHOST. This characteristic was originally promulgated and used by Robson and Reed (1999) in the development of an equation for predicting the median flood (QMED) from catchment characteristics/descriptors. HOSTRES indicates the extent to which BFIHOST is higher or lower than that anticipated from SPRHOST, and as such is uncorrelated with both BFIHOST and SPRHOST.

**Table 5.7 Fractions of HOST classes within the UK and within gauged catchments**

| Class | Fractional Extent in UK | % of catchments class occurs | mean fractional extent of occurrence |
|---|---|---|---|
| HOST1 | 4.5 | 14.8 | 32.2 |
| HOST2 | 1.7 | 14.5 | 12.5 |
| HOST3 | 3.5 | 19.2 | 6.9 |
| HOST4 | 5.4 | 43.5 | 8.1 |
| HOST5 | 1.8 | 61.3 | 5.8 |
| HOST6 | 0.8 | 51.3 | 5.3 |
| HOST7 | 1.6 | 51.5 | 1.3 |
| HOST8 | 3.8 | 61.3 | 0.9 |
| HOST9 | 2.9 | 58.8 | 1.3 |
| HOST10 | 0.6 | 64.6 | 2.1 |
| HOST11 | 0.5 | 10.0 | 1.5 |
| HOST12 | 0.9 | 38.2 | 5.2 |
| HOST13 | 0.7 | 37.0 | 2.3 |
| HOST14 | 10.4 | 23.7 | 3.1 |
| HOST15 | 0.4 | 68.8 | 19.6 |
| HOST16 | 10.7 | 23.7 | 3.6 |
| HOST17 | 5.7 | 61.8 | 22.3 |
| HOST18 | 2.3 | 57.4 | 9.6 |
| HOST19 | 0.7 | 35.9 | 5.0 |
| HOST20 | 4.2 | 15.6 | 4.6 |
| HOST21 | 1.2 | 29.0 | 12.3 |
| HOST22 | 1.4 | 38.2 | 4.2 |
| HOST23 | 14.6 | 14.2 | 8.4 |
| HOST24 | 3.8 | 78.8 | 16.9 |
| HOST25 | 2.6 | 18.1 | 21.5 |
| HOST26 | 0.9 | 43.7 | 11.3 |
| HOST27 | 0.6 | 17.0 | 3.4 |
| HOST28 | 8.5 | 7.0 | 8.6 |
| HOST29 | 2.2 | 61.6 | 14.7 |
| HOST30 | 0.6 | 34.0 | 0.8 |

The Institute of Terrestrial Ecology Landcover Map of Great Britain

The ITE Remote Sensing Unit derived the land cover map from 1990 and 1991 LANSAT satellite images (Fuller *et al*, 1994). The derived classes represent an aggregation of many subclasses: for example, wheat, barley and oilseed rape would be subclasses of the 'arable' class. The objective when deriving the classification was to derive a set of classes that are considered ecologically meaningful, consistently recognisable from the selected imagery, and realistic in terms of their likely accuracy. Classes were only mapped if they were greater than 0.125 ha in extent.

The classes distinguish lowland and upland categories that are similar, for example lowland heather and upland dwarf shrub. The version of the map used for the study was the 1km resolution full 25-class map in which the fractional extents of cover classes within each grid cell are held.

One concern when using land cover data relates to whether the land cover classification for a catchment is representative of the cover across the period of record for which there is flow data. If not (as is likely to be the case), serious consideration needs to be given to whether the data is included in the analysis. In practice the Landcover map was not found to be a particularly useful characteristic other than providing a mechanism for setting interception parameters for MODA as described in Chapter 6. The 25 classes are also presented in Chapter 6.

## 5.3 SUMMARY OF CATCHMENT CHARACTERISTICS AND DESCRIPTORS

The sampling of the spatial variability of HOST within the set of catchments selected for this study is discussed within the preceding section. The median and 95 percentile limits (95% of catchments have values between these limits) for the values of the continuous (rather than class based) characteristics/descriptors across the candidate catchments is presented in Table 5.8.

**Table 5.8 Summary statistics for continuous catchment characteristics and descriptors**

| Percentile | LDP | DPLBAR | DPLCV | ALTBAR | DPSBAR | ASPBAR | ASPVAR |
|---|---|---|---|---|---|---|---|
| 95% u.l. | 102.79 | 54.85 | 0.54 | 556.63 | 298.09 | 351.72 | 0.48 |
| Median | 24.31 | 12.99 | 0.43 | 228.10 | 107.22 | 133.30 | 0.19 |
| 95% l.l | 3.59 | 1.89 | 0.35 | 47.73 | 19.71 | 4.14 | 0.05 |

| | AREA | SAAR(61-90) | PE(61-90) | PP | BFIHOST | SPRHOST | HOSTRES |
|---|---|---|---|---|---|---|---|
| 95% u.l. | 1509.67 | 2511.75 | 622.72 | 0.28 | 0.89 | 55.62 | 0.09 |
| Median | 110.25 | 1064.00 | 511.95 | 0.03 | 0.45 | 40.50 | 0.01 |
| 95% l.l | 3.19 | 594.93 | 422.76 | 0.00 | 0.25 | 10.63 | -0.11 |

It is useful to review the sampling of catchment characteristics and some of the catchment descriptors (which are a function of the specific catchments) in the context of the spatial variability in the United Kingdom.

The figures for catchment AREA and LDP illustrate that the sample of catchments is skewed towards smaller catchments. This is a function of the necessity for selecting natural catchments, which tend to be headwater catchments. This is advantageous from the perspective that smaller catchments also tend to be more homogenous in terms of hydrogeology and also means that it is not necessary to address channel routing considerations. This does leave the question to be answered on how a regionalised model should be applied to large catchments. This question is returned to within Chapter 10. The relatively high values observed for ALTBAR demonstrate that headwater catchments tend to be at higher altitudes than their lowland larger counterparts. This is also demonstrated to a lesser extent in the values for DPSBAR where the median mean catchment slope is approximately 10%. There appears to be no bias in catchment aspect (ASPBAR) and catchments are not highly elongated (DPLCV lower limit of 0.35).

Considering the climatological characteristics, the range of PE values is commensurate with the range of PE across the United Kingdom (Chapter 2). The 95% upper limit of 2500mmyr$^{-1}$ for SAAR is considerably lower than the range observed across the UK and illustrates the point that very high rainfall catchments tend to be at high altitude and are not gauged. The reasons for this are access and the lack of an operational requirement for routine monitoring of flows in upland catchments. The range of PP values demonstrates that there is the potential for summer soil moisture deficits to commonly occur in approximately 50% of the study catchments.

# 6 The model calibration scheme and an evaluation of model behaviour

## 6.1 INTRODUCTION

Model parameters may be classed as physical parameters and process parameters (Sorooshian and Gupta, 1995). The former class represents measurable catchment properties, the commonest example of which is the catchment area. Process parameters are those that are not directly measurable. Common examples include soil storage capacities and time constants for routing reservoirs. The process of calibration seeks to identify a combination of parameters such that the simulated stream flow response to precipitation closely matches that which is observed. The common practice for identifying a suitable parameter set, or vector, for the model consists of automatic calibration coupled with visual inspection of the hydrographs of simulated and observed stream flow and manual intervention in the selection of model parameters. The model calibration scheme for this research was required to automatically calibrate the models on a large number of catchments in a consistent manner, such that the values for parameters, within the identified parameter vector, for each catchment should:

- have physical meaning, that the parameters can potentially be expressed as a function of the physiographic and climatic characteristics differentiating between catchment types;
- provide a good simulation, from a water resources perspective, of stream flow.

Due to large number of catchment involved, it was not a logistically feasible proposition to calibrate the models using large-scale manual intervention. Furthermore, manual intervention involves the making of subjective decisions. As discussed in Chapter 1, a good simulation from a water resources perspective requires an assessment of many different aspects of the simulated time series. Additionally, the calibrated model must also be stable when it is applied to an independent evaluation period.

The objective of the work described within this chapter was to develop a scheme that could automatically implement the sort of decisions that an experienced hydrologist would make as a set of fuzzy rules. As will be discussed, the scheme developed is a constrained random walk scheme that utilises a number of objective functions, allows for trade off between different aspects of model fit and recognises that the input stream flow data have an associated uncertainty. Within a catchment, the scheme identifies a large number of parameter sets, or vectors, that can be considered equally likely given these constraints. The final selection of a model parameter vector (to represent the catchment with the regionalisation analysis) is made on the basis of the performance of the calibrated model for each vector over the calibration period and the stability of the model when applied to an independent evaluation period.

The issues associated with model calibration and the rationale for the model calibration scheme are discussed within section 6.2. The selection of appropriate objective functions and additional goodness of fit measures is presented within Section 6.3. The calibration scheme is presented within Section 6.4.

The calibration scheme, as it realises a large number of equally likely parameter vectors, was also used to investigate parameter identifiability and covariance issues within the model structures across a range of catchment types. The issues raised by these investigations are discussed within Section 6.5. These issues are illustrated using four contrasting case study catchments. The rationale underlying the development of MODB is also discussed within this section. The objectives of the parameter identifiability research described in Section 6.5 were:

- to define a generalised, consistent approach for calibrating the models across a wide range of catchment conditions;
- aid the definition of model structures that minimise the identifiability and covariance issues associated with parameter redundancy;
- aid the definition of a consistent procedure for identifying an optimal parameter vector, from many such vectors, that provides a good, stable simulation of stream flow from a water resources perspective.

Optimal in this context does not imply identifying the vector that corresponds to the global minimum of the objective function surface (discussed within the next section) - which may, in part, be an artefact of the calibration period selected and the errors associated with the input data for the period. Rather, the approach seeks to identify a parameter vector that is stable, and that can not be said to be significantly different from the optimal solution, in the context of the uncertainties associated with the input data. The procedure derived for the final selection of a model parameter vector for a catchment is presented within Section 6.6.

## 6.2 ISSUES WITHIN MODEL CALIBRATION

Automatic calibration requires the definition of a measurable quantity, termed the objective function, that is a measure of how closely the simulated time series matches the observed time series. These functions are commonly structured so that the best simulation is represented by a minimum value. The minimum value of the objective function will be a function of the model structure, the parameter values and the input data for the calibration period, including the errors within the input data. The simulation errors that can be introduced through errors in climatic input data are discussed within Chapter 2. An automatic calibration scheme consists of two basic components.

- The objective function(s), which describes the quality of fit between the observed and the simulated stream flow derived using a particular parameter vector (or set of parameters).
- An automatic search algorithm, including stop criteria, for searching the objective function response surface, that is described by the feasible parameter space for the model, with the objective of identifying a minimum, or optimal value of the objective function.

In addition to these basic components, there is also the potential *a priori* determination of model parameters and the sensitivity of the model response to the choice of model parameters. Considering one objective function, the allowed ranges for the n model parameters therefore describe an n-dimensional objective function space. There is the concept of a global minimum within the objective function space. This minimum is described by the parameter set that gives the lowest, attainable value of the selected objective function. Model calibration is, normally, directed at identifying the parameter

vector describing the location of this global minimum that represents the unique solution.

It may not be possible, in practice to identify a parameter vector describing the location of the unique, or global, minimum solution. A major problem associated with identifying the location of a global minimum is that local minima may operate on different scales. The objective function surface may contain several large areas of attraction (i.e. areas where the objective function appears to be converging to a minimum), and within these areas there may be discrete local minima representing "pockets" within the larger areas (Duan *et al*, 1992). The global minimum is the deepest of these pockets that contains the minimum value of the objective function. The problem of identifying the global minimum generally scales with the complexity of the model, and hence the number of model parameters. Jakeman and Hornberger (1993) suggest that a maximum of about six model parameters can be identified from stream flow data for temperate catchments. Beven (1989) suggests that three to five parameters should be sufficient to describe most aspects of an observed hydrograph.

Research into the identification of the global minimum for a model over a specified calibration period has been the subject of many papers, for example: Ibbit, (1970), Johnston and Pilgrim (1976), Soorooshian and Gupta, (1983) and Duan *et al*, (1992). Furthermore, it has been identified that the regions of the objective function surface where "optimal" solutions might lie can vary markedly between selected calibration periods (e.g. Yapo *et al*, 1996) due to errors in the data and/or in-adequate representation of all hydrological process that the model has to simulate. Beven (1993) disputes the existence of the global minimum as a general concept, recognising that there may be many different parameter vectors providing indistinguishable minimum values for the objective function, he terms this the *equi-finality* of solution.

Search algorithms used within automatic calibration schemes can be classed as either local or global search schemes. Examples of common local search schemes include the Rosenbrock method (Rosenbrock, 1960) and the Simplex method (Nelder and Mead, 1965). The stop point for all search algorithms is either a minimum value for the objective function, or a convergence criterion specifying under what conditions can it be assumed that no further improvements will be identified. Local search algorithms are structured to

follow a direction of an improving objective function. The path taken is a function of the initial starting point, and the methods are susceptible to becoming "trapped" in a local minimum.

Global search strategies are structured to search objective function surfaces that are multi-modal (i.e. contain local minima). The strategies fall in to two classes: random walks and Multi-Start Local (MSL) searches. Multi-start algorithms are basically a large number of independent runs of a local search procedure using randomly selected starting point parameter vectors, selected from a specified parameter space for the model (e.g. Johnston and Pilgrim, 1976). Duan *et al* (1993) extended a simplex-based MSL concept by sharing knowledge between individual parallel searches. This method, called the Shuffled Complex Evolution Algorithm (SCE-UA) has been demonstrated to be to be superior to both local and MSL search algorithms (Duan *et al*, 1993, Franchini *et al*, 1998). The other class of global search algorithms are based on random walks of the feasible parameter space using probability distributions (usually uniform) to describe the frequency of occurrence of parameters across the feasible space. The pure random search assumes that the optimal parameter vector may lie anywhere within the feasible space. The Adaptive Random Search (ARS) (Masri *et al*, 1978 and Pronzato *et al*, 1984) uses knowledge about prior randomly obtained parameters vectors to guide the search.

A more recent development in random walk algorithms is the Generalised Likelihood Uncertainty Estimation (GLUE) method (Binley *et al*, 1991; Beven & Binley, 1992; Beven, 1993). A large number of simulations are made by randomly selecting parameters from a feasible parameter space for the model. Where two parameters are demonstrably covariant, the feasible space for these parameters can be defined by a conditional bi-variate density function. Each simulation is assigned a "likelihood" weight in the range [0,1]. This is based on a subjective "likelihood" function. This function may be based on statistical considerations and/or on a subjective measure of what is a desirable simulation in the context of model output. The pool of simulations can, at any point, be used to derive a probability density function based on the values of the likelihood function. From this, non-parametric uncertainty intervals (analogous to the confidence interval from classical statistics) can be derived for judging whether simulations are acceptable model realisations on the basis of the likelihood function. As knowledge of the model behaviour improves, the

likelihood function can be further constrained to reduce the uncertainty bounds. The method does require a definition of a feasible parameter space and knowledge about covariance issues, however these can be further refined as the knowledge base of model behaviour expands.

Calibration errors can be introduced through errors in input data, the use of incorrectly specified objective functions and/or the failure to identify the global minimum, or at least a stable local minimum, within the objective function space. Careful selection and quality control of input data and selection of appropriate objective functions can reduce these errors. The presence of multiple minima is normally associated with structural model errors (Sorooshian and Gupta, 1985). Structural issues can commonly arise from models being over-parameterised in quest for physically realism, leading to parameter covariance. Parameter covariance is when the same quality of fit may be obtained for a wide range of parameter combinations for two or more parameters. The issue of parameter covariance can be a primary cause of failure to identify a global or acceptable minimum.

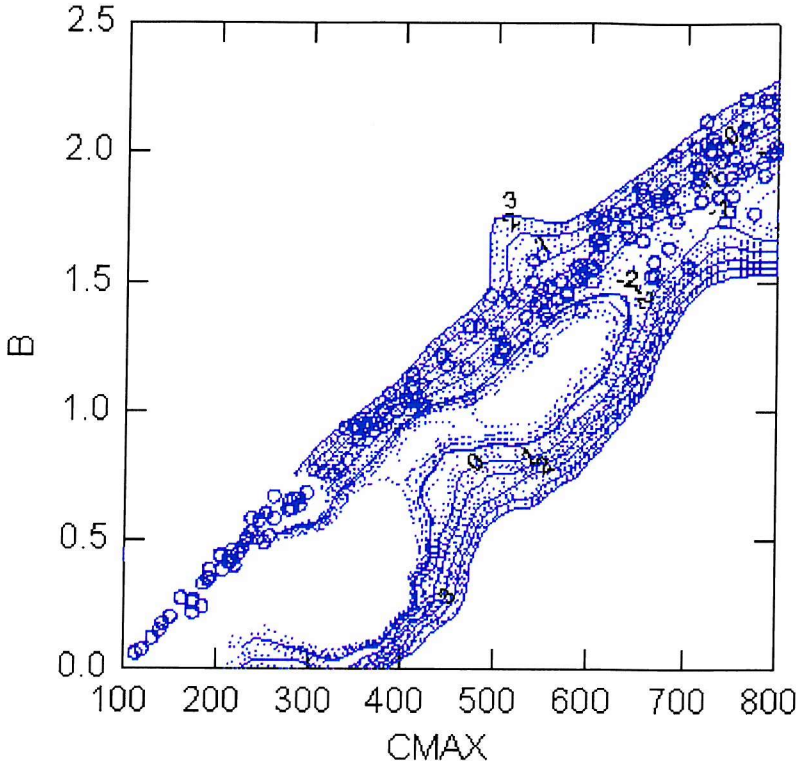
Two examples of parameter covariance are presented in Figure 6.1a and Figure 6.1b. These are drawn from this study and relate to the calibration of MODA on the River Pinn at Uxbridge (see Section 6.5 for details of this catchment). These show contoured plots of the value of the simulation BIAS objective function (discussed in the subsequent section) for combinations of loss module parameters for MODA. Figure 6.1a presents BIAS contours for combinations of Cmax and b. Contours of simulation BIAS for Cmax and Kg are presented within Figure 6.1b for the same simulation runs. The values of simulation BIAS vary between –3 and +3%. The ideal value for simulation BIAS is obviously zero. The contour plot for BIAS, as a function of Cmax and b describes a long valley encompassing a wide range of values for Cmax and b. The valley bottom has a BIAS value of –3 and the optimal value of zero lies on the flanks of the valley. This is the classic example of covariance, where the same value of an objective function is realised for a wide range of covariant parameters, and is an example of potential structural defect within the model. Cmax and Kg are not significantly covariant. In the case of Cmax and Kg, the BIAS contours describe a number of areas of attraction in which the ideal minimum BIAS of zero is to be found. This is an example of the occurrence of local minima.

On this basis of this data there appears to be major structural defects within the model. However, considering the [-3,3] data range for simulation BIAS there is a strong argument that this range lies within the measurement error for stream flow measurement. Thus all of these simulation runs could be considered as being equally good in the context of closing an acceptable water balance for this catchment.

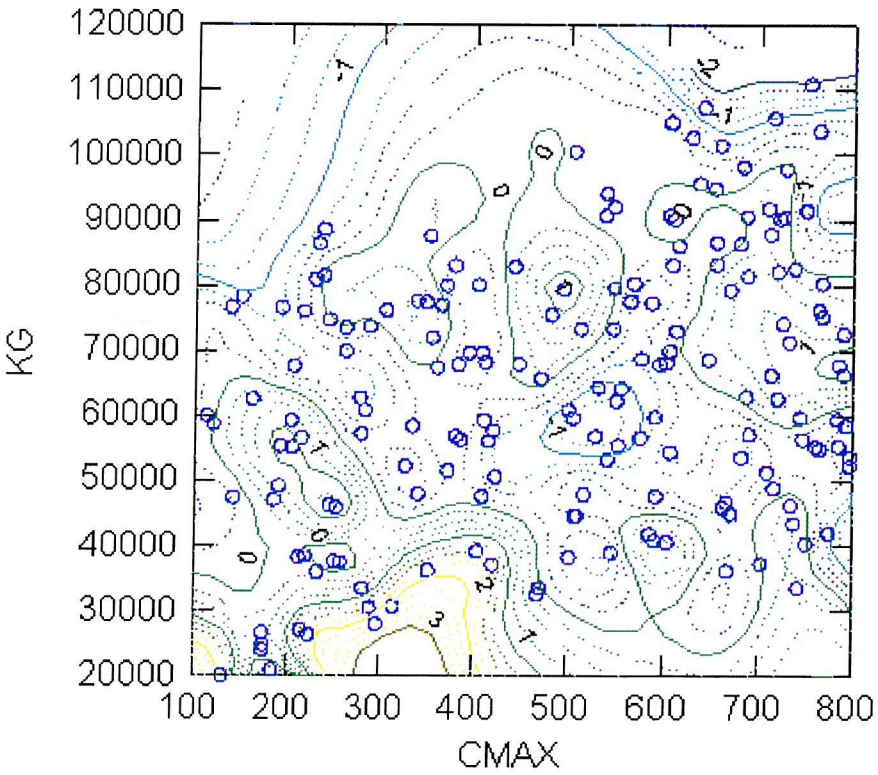
In practice, it is generally accepted that two or more model parameters will exhibit some degree of covariance, and that the parameter vector identifying the theoretical global minimum may not be distinguishable from many equally likely parameter vectors, particularly when considering the accuracy of calibration input data.

The issue of equi-finality is of importance in regionalisation for two reasons. Firstly the optimal fit obtained over the calibration period may not be stable when the model is applied, with that parameter vector, to an independent evaluation period. Secondly, if for a particular model similar optimal solutions, and hence parameter vectors, cannot be identified for similar catchment types, then it is unlikely that research into the relationships between those parameters and the characteristics of the catchment is going to be successful.

Even if a global minimum can be identified for an objective function, the objective function itself may not adequately summarise all of the facets of the flow regime that are of interest for the application. In the context of water resources the practitioner is interesting in a number of facets of the stream flow simulation. These include the closure of an adequate water balance as represented by model BIAS and the correct simulation of recessions. This requires the use of more than one objective function. In the automatic calibration schemes described, the model fit is generally optimised according to one best compromise, objective function and subsequently other aspects of the model fit are evaluated using a range of secondary measures of fit. The primary reason for this is that differing objective functions may, in the worst case, be orthogonal to one another within the parameter space.




**(a)**




(b)

**Figure 6.1 Examples of parameter covariance for model BIAS.**

With manual intervention, subjective decisions about the trade off between different measures of fit and judgements about the quality of fit, in the context of the likely accuracy of the input data can be made. These types of decision cannot be addressed within the type of automatic search algorithms described. In the context of the large number of catchments used within this study it was not a practical option to manually intervene in a standard automatic calibration scheme to make these types of decision.

The automatic calibration scheme developed for the study is a random walk scheme that seeks to mimic these human decisions. There are some parallels with GLUE in that it is recognised within the scheme that many parameter vectors may be considered as being equally likely in representing a best fit vector. However, where GLUE does not define a specific parameter set, or vector, as an optimal solution, the current scheme does identify a feasible, best compromise vector for a catchment. The scheme uses a number of objective functions to measure the quality of a solution. In identifying potential "best fit" parameter vectors a trade off between objective functions, subject to the uncertainty within the input data is allowed. The scheme identifies a large number of potential "best fit" parameter vectors that can be considered as being equally likely when the uncertainty of model input data and the requirements of differing objective functions is considered. The parameter vector that is ultimately selected for the regionalisation studies is the vector that represents the best compromise between the quality of the model fit over the calibration period and the stability of the model fit when the vector is applied to a evaluation period. This selection builds on the discrete model evaluation scheme developed in Chapter 3 to evaluate model performance. This scheme was also used to investigate the degree of parameter covariance within different catchment types with the objective of understanding the model response within different catchment types leading to an identification of a model structure that had minimal inherent parameter covariance.

The scheme developed for this study uses four objective functions for identifying equi-likely parameter vectors and an additional evaluation function describing the "goodness" of fit at low flows. The distinction between the function types is that the objective functions are used to select model parameter vectors within the calibration phase. The choice of objective functions and goodness of fit measures is discussed in Section 6.3.

## 6.3 SELECTION OF OBJECTIVE AND GOODNESS OF FUNCTIONS

The choice of objective function for the study was based on the requirements of a good simulation for water resources purposes. For water resources assessments it is important to ensure that mass is conserved, i.e. mean flow is accurately simulated, and that low flows (taken as flows below the mean) and low flow extremes are modelled correctly. The fit at high flows is less important, as long as mass is conserved. The issue of whether a selected objective function accounts for the time series element of flow, or whether it is distribution based is also influenced by the intended use of the model. If the results are to be summarised statistically, then the accuracy of the sequencing may be less important. However, if the sequencing is correct this gives much greater confidence to the validity of the model, and thus there is a strong case for it always to be considered.

Whilst the evaluation of particular objective functions is not a widely researched area in hydrology, examples of such evaluations are given in Thian et al (1997), Houghton-Carr (1999) and Servat & Dezetter (1991). Further examples of particular objective functions can be found within the literature on model inter-comparisons; for example Hughes (1995) and Chiew *et al* (1993).

The most widely used, quantitative objective functions are ones that draw from linear regression analysis. The most common and simplest formulations of these functions include the sum of squares error (S.S.E), the mean of the sum of squares error (M.S.E.) and the root mean sum of squares error (R.M.S.E.). The R.M.S.E. function is the more familiar standard error, or expression of unexplained variance, from linear regression. The formulation of the S.S.E is presented here for discussion purposes:

$$SSE = \sum_{i=1}^{n} (\hat{q}_i - q_i)^2 \quad , \qquad \textbf{(6.1)}$$

where:

$q_i$ = observed flow in time step, i;

$\hat{q}_i$ = simulated flow in time step, i;

n = number of time steps within the modelled period.

As stated, sum of squares based objective functions arise out of linear regression, the basic principles of which are:

- the underlying relationship between the dependent variable and the independent variable is linear, or can be made so by a suitable transformation;
- the independent variables are without error;
- the errors in the dependent variable (in this case predicted stream flows) are normally distributed with zero mean, and furthermore they are independent of one another;
- the conditional variance of the dependent variable should be constant across all values of the independent variable;
- the independent variable data set should be uncorrelated, and therefore independent.

Measured daily mean flow data violate the independence test for the independent variable in that the observed flow within a time step is strongly related to the flow in the previous time step. This is demonstrated in the partial auto correlation plots for the Babingley Brook (used within the model evaluation studies) and the Teifi given in Figure 6.2. The Babingley Brook (33054) is predominantly unconfined Chalk, and thus the flow regime is heavily dominated by groundwater discharge. The BFI for the catchment is 0.98. In contrast the Teifi at Glan Teifi (62001) has a catchment area of 894 $km^2$ and Ordovician and Silurian deposits dominate the geology of the catchment. The BFI of the catchment is 0.53 indicating a more responsive flow regime.

Partial auto-correlation plots identify the correlation between the ith and (i-n)th element within a time series, but remove the influence of the correlation between the ith and the (i-(n-1))th elements on the correlation between the ith and the (i-n))th elements of the time series. The plots demonstrate that a high degree of correlation exists between adjacent elements within the time series of daily mean flows in both permeable high storage systems, such as the Babingley, and in low storage impermeable catchment, such as the Teifi. The same constraints apply to the dependent variable, simulated stream flow, the consequence of which will be that errors are not independent.

```
NUMBER OF CASES = 6016
MEAN OF SERIES =       0.567
STANDARD DEVIATION OF SERIES =       0.240

PLOT OF PARTIAL AUTOCORRELATIONS

 LAG    CORR      SE -1.0  -.8  -.6  -.4  -.2   .0   .2   .4   .6   .8  1.0
                     +----+----+----+----+----+----+----+----+----+----+----+
   1    .989    .013                           |XXXXXXXXXXXXXXXXXXXXXXXXX
   2   -.012    .013                           |
   3    .175    .013                           |XXXX
   4    .065    .013                           |X
   5    .062    .013                           |X
   6    .027    .013                           |
   7    .008    .013                           |
   8   -.023    .013                           |
   9   -.014    .013                           |
  10   -.021    .013                           |
  11    .016    .013                           |
  12    .033    .013                           |
  13    .019    .013                           |
  14    .006    .013                           |
  15   -.050    .013                          X|
```

**(a) Babingley Brook**

```
NUMBER OF CASES = 7305
MEAN OF SERIES =      28.041
STANDARD DEVIATION OF SERIES =      30.288

PLOT OF PARTIAL AUTOCORRELATIONS

 LAG    CORR      SE -1.0  -.8  -.6  -.4  -.2   .0   .2   .4   .6   .8  1.0
                     +----+----+----+----+----+----+----+----+----+----+----+
   1    .910    .012                           |XXXXXXXXXXXXXXXXXXXXX
   2   -.199    .012                      XXXX|
   3    .132    .012                           |XXX
   4    .020    .012                           |
   5    .051    .012                           |X
   6    .030    .012                           |
   7    .035    .012                           |
   8    .030    .012                           |
   9    .029    .012                           |
  10    .005    .012                           |
  11    .025    .012                           |
  12    .017    .012                           |
  13   -.015    .012                           |
  14    .013    .012                           |
  15    .035    .012                           |
```

**(b) Teifi**

**Figure 6.2 Partial auto-correlation plots for the Babingley Brook and the Teifi.**

The errors in simulated stream flow data are also not of a constant variance. The non-constant behaviour of variance is a consequence of the strong non-linearity of the rate of change (the second derivative) of river flows over the range of the flow regime. Furthermore, within the distribution of daily stream flows, the low and high flow extremes are poorly sampled, yet these extremes (particularly the high flows) have a strong influence on the value of the S.S.E. The S.S.E. for observed and simulated stream flow in a catchment is likely to be biased toward minimising the difference between a relatively small number of high flows. Furthermore, the error structure may be complicated due to the

correlation between flows within adjacent time steps - although this is less of a problem at high flows than low flows. This can be ameliorated; either by minimising the sum of squares for the logarithms of flows or by applying a time dependent weighting factor related to the magnitude of flow within the time step. This weighting factor is commonly structured such that it transforms errors to a constant variance using a maximum likelihood estimator (Sorooshian & Gupta, 1995). Although the influence of poorly sampled high flow events on the sum of squares can be minimised by taking logarithms, this does not address the associated problem at low flows. The influence of low flows in the sum of squares is affected by two factors; the low frequency of occurrence and the relatively small differences observed at low flows.

Other commonly used "sum of squares" objective functions are the measure of explained variance from linear regression, $R^2$:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{q}_i - \bar{q})^2}{\sum_{i=1}^{n} (\hat{q}_i - \bar{q}_i)^2 + \sum_{i=1}^{n} (\hat{q}_i - q_i)^2}, \qquad \textbf{(6.2)}$$

where $\bar{q}$ is the mean of the observed flows, and the Nash-Sutcliffe efficiency criterion (Nash and Sutcliffe (1970):

$$\text{Efficiency} = 1 - \frac{\sum_{i=1}^{n} (\hat{q}_i - q_i)^2}{\sum_{i=1}^{n} (q_i - \bar{q})^2}. \qquad \textbf{(6.3)}$$

The sum of squared differences in the numerator of the efficiency criterion can be considered as the residual variation, or unexplained variation, whilst that in the denominator is the total variation. The efficiency tends towards one as the simulated flow tends to the observed. If the efficiency is negative it indicates that a better model fit would be obtained if the simulated flows were replaced by the observed mean flow. It is worth noting the close similarity of the Nash-Sutcliffe efficiency criterion to the definition of $R^2$ given in equation 6.2.

If the least square relationships between observed and simulated flows is linear and observed and simulated flows comply with the assumptions of linear regression then

$$\sum_{i=1}^{n}(q_i-\overline{q})^2=\sum_{i=1}^{n}(\hat{q}_i-q_i)^2+\sum_{i=1}^{n}(\hat{q}_i-\overline{q}_i)^2 . \qquad \textbf{(6.4)}$$

Substituting equation 6.4 for the total variance within equation 6.3 yields the definition of $R^2$ given in equation 6.2. As the efficiency is not constrained to lie between zero and one, it is quite common to obtain a simulated time series that has a reasonable $R^2$ value but a negative value for efficiency. This is a consequence of the limitations of assuming the data complies with the assumptions of linear regression.

The measures $R^2$ and efficiency have the advantage that they are dimensionless and thus can be used for inter-catchment comparisons. It is important to recognise that the sensitivity of the measures to model errors is a function of the inherent variability of the flow regime. For a given quality of fit, one would expect to observe higher (better) values of these measures in impermeable, flashy catchments than in more permeable catchments with an inherently lower variability. Nielsen & Hansen (1973) recognised this problem with respect to correlation coefficient dependencies on mean seasonal variations in observed flows and proposed an alternative correlation measure. In this measure the $R^2$ between observed and simulated flows for each calendar day across the n years of data being modelled is calculated and the average of the resultant 365 coefficient values is taken. The limitation of this approach is that there is the inherent assumption that the calendar day is a hydrological significant time unit.

The objective functions used in the calibration scheme were:

- simulation bias (BIAS);
- the modulus of the difference between observed and simulated Base Flow Index (DBFI);
- Nash Sutcliffe efficiency criterion (EFF);
- the mean sum of squared differences between observed and simulated flows for observed flows that less than the flow that is equalled or exceeded for 67% of the time (i.e. periods within the observed stream flow when recessions occur): LF_OBJ.

The simulation BIAS is a measure of whether the model is closing a water balance or not (and hence modelling mean flow correctly), and is calculated within the study using

$$\text{BIAS} = \left( \frac{\bar{\hat{q}}}{\bar{q}} - 1 \right) \text{x} 100 \ . \qquad \textbf{(6.5)}$$

In this configuration a negative value indicates an underestimate of mean flow, and a positive value an over estimate of mean flow.

The BFI is the ratio of the volume of water derived from the slow flow component of a hydrograph to the total volume of water over a specified period of record, and is a good general measure of hydrological response. In this case it was used as a measure of the realism of the partitioning of effective rainfall within a model run. The BFI for a catchment was calculated from the gauged flow data using a hydrograph separation algorithm described in Natural Environment Research Council (1980), and illustrated in Figure 6.3.

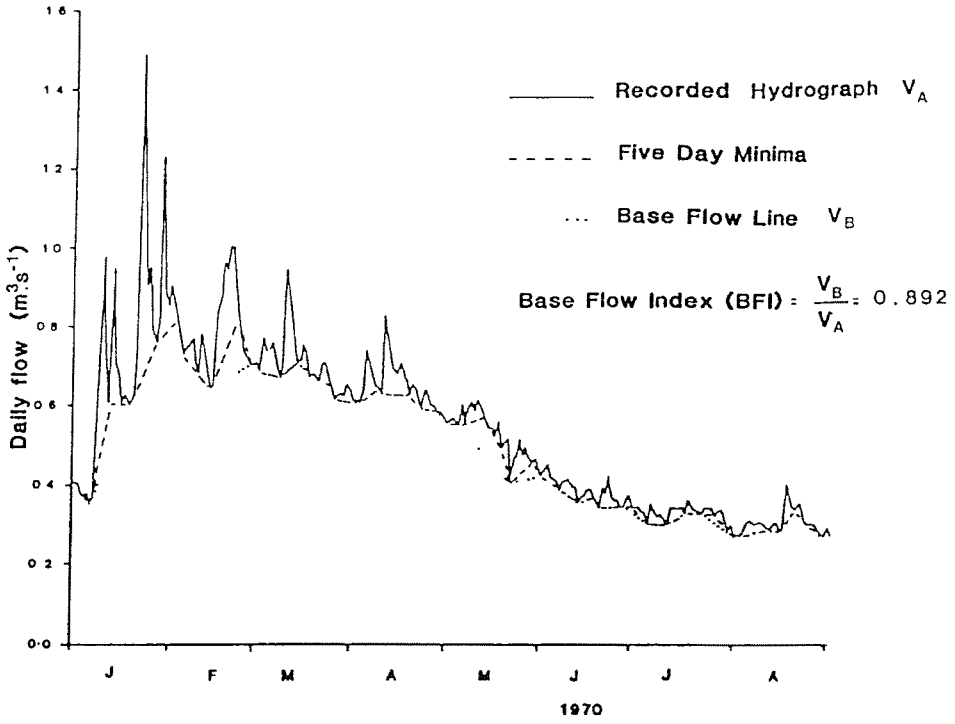



**Figure 6.3 Summary of the base flow separation algorithm for observed data** ***(Source: Gustard et al, 1992).***

The simulated BFI for a model run was calculated as the ratio of the volume of the effective rainfall routed through the slow flow response path, $U_b$ of the model to the total volume of effective rainfall over the simulation period, $U_b+U_s$:

$$BFI_s = \frac{\sum_{i=1}^{n} Ub_i}{\sum_{i=1}^{n} (Us_i + Ub_i)} \quad . \qquad \textbf{(6.6)}$$

Simulation trials demonstrated that, even for permeable catchments with large time constants for the slow flow routing reservoir, this ratio is close to the ratio of simulated base flow volumes (output from the slow flow routing reservoir) to total simulated volume for simulation periods of greater than three years. This demonstrates that the change in storage within the routing reservoirs is negligible compared to the total volume of water for simulation periods of greater than two to three years. Effective rainfall volumes, rather than

model output volumes were used to estimate the simulated base flow as the measure was used as part of the identification of the realistic parameter space for loss module parameters during the calibration procedure as described in the next section. For the same reasons the trials were also used to demonstrate that the calculation of the BIAS statistic was insensitive to whether the mean simulated effective rainfall, or the mean simulated stream flow was used in the calculation of the BIAS statistics.

The LF_OBJ objective function reflects the requirement to model recession periods, i.e. periods of low flow, accurately. This objective function was used in the calibration of the routing module parameters and the final calibration of the loss module parameters. The use of the Nash-Sutcliffe Efficiency Criterion reflects the requirement to model the general variation in the time series accurately. In addition to the formal objective functions described the mean bias error at the observed Q95 flow (BEQ95) was used as an additional goodness of fit measure. This measure was calculated by extracting from the observed time series all flows that had exceedence percentiles lying between limits of Q94.49 and Q95.49 (inclusive). The corresponding simulated flow values were also extracted for the selected time steps and the value of BEQ95 calculated using the same formulation as used for the BIAS statistic. This additional statistic was used to provide information on whether there is a consistent bias in a simulation at a key low flow percentile. This is useful to distinguish between biased and unbiased values of LF_OBJ.

## 6.4 THE CALIBRATION SCHEME

### 6.4.1 The calibration algorithm

The calibration scheme developed is a three-stage approach based upon a Monte Carlo sampling strategy of the feasible parameter space. The scheme initially identifies a realistic parameter space prior to identifying parameter vectors that result in equally acceptable simulations of stream flow over the calibration period. A feasible parameter space is one in which the parameter ranges are suitable for all catchments, and is defined by upper and lower limits for parameter values. The realistic parameter space is defined as a catchment specific subset of the feasible parameter space that contains the significant areas of attraction within the objective function surface(s). For the majority of catchments, in the

context of this scheme, the realistic parameter space is a reduced space that contains those regions of the feasible space where the model BIAS is less than ±3% and the simulated BFI is within ±0.15 of the BFI estimate from the observed data.

The BFI criterion ensures that only regions of the feasible space in which the models realistically partition effective rainfall between quick and slow-flow routing paths are selected. The BIAS criterion of ±3% represents an estimate for what a good quality, well maintained gauging station structure should be able to achieve in terms of hydrometric accuracy at a confidence level of approximately one standard deviation. This is a stringent criterion. In practice, if the criterion could not be met within a catchment, it was relaxed to ±5% to boost the sample size of successful calibrations. Assuming a gaussian measurement error distribution for gauged flow, this relaxed constraint is tighter than a 95% confidence interval, which would be approximately ±6%.

The calibration scheme consists of three stages. In the first stage the loss module is calibrated to identify a realistic parameter space for the loss module, starting from initial upper and lower feasible limits for the individual parameters. In the second stage realistic optimal, or target values for the LF_OBJ and EFF objective functions are estimated, given:

- the constraints of input data accuracy;
- the (sometimes) conflicting regime requirements for optimising the LF_OBJ and EFF functions;
- the necessity for closing an effective water balance and maintaining a realistic partitioning of effective rainfall, as defined by the end point criteria for BIAS and the difference in simulated and observed BFI (DBFI).

In the third stage these target values are then used to identify 300 equally likely, or valid parameter vectors for the model that, given the accuracy of the input data and the requirement from the simulated flows can be considered as giving equally good stream flow simulations. A fuzzy logic approach is adopted for testing whether a particular shot is valid.

A shot is accepted as being valid if:

- the shot values of BIAS and DBFI are less than or equal to the end point criteria from the loss module,
- AND the value of one of the routing objective functions (LF_OBJ or EFF) is equal to or better than the reasonable target limit identified within the second stage,
- AND the value of the second objective function is within +20% of the identified reasonable target limit for the second objective function.

The setting of a fuzzy limit allows for a trade off between the quality of fit over the whole time series and that at low flows. This is an attempt to mimic the type of trade off a modeller will make when evaluating the quality of a model fit through visual inspection. The 300 parameter values obtained in this way can be considered as being equally likely, given the various requirements for accurate simulation, as represented by the objective functions, and the inaccuracies within the input data.

The method requires a definition of the parameter space for each parameter (discussed in the subsequent sub-section) and the definition of a probability distribution function for the parameter space from which to draw samples. This involves a degree of subjectivity. As will be discussed, the setting of a feasible limit for Kg in MODA was difficult. To address this, the limits for Kg were programmed in as "fuzzy" limits. During the first three iteration loops for the loss module the feasible limits for Kg are not fixed. If a valid shot has a value of Kg that occurs within 5% of either limit, the space for Kg is moved by 10% in that direction. After loop three it was empirically demonstrated, over a wide range of catchment types, that the scheme had found the feasible space containing the significant areas of attraction within the Kg parameter space. The same approach was adopted for the initial setting of feasible limits for the time constant for the quick-flow routing reservoir, K1.The detail of the calibration scheme is presented within Appendix B.1.

This heuristic, Monte–Carlo based calibration procedure differs from conventional search algorithms, in that the selection of individual valid parameter vectors are independent of one another. The approach is computationally intensive, despite the use of the first stage to reduce computational overheads. It does mean however, that the parameter space is well

sampled to provide a range of valid parameter vectors from which to make the final selection of a parameter vector for use within the parameter regionalisation studies, described in Chapter 7. The final choice of parameter vector is described in section 6.5. The calibration period used for each catchment was 10 years. This period is much longer than that commonly used in model calibration. Sorooshian and Gupta (1995) suggest that between 2 and 3 years of representative flow data should be sufficient. A 10-year period was used to ensure that the underlying population of stream flows was being adequately sampled. Where possible, the calibration period corresponded to the latest 10 year period encompassing the 1990-1992 period over which the land cover data set (used as one of the catchment characteristics) was developed (Chapter 5). A two-year model "warm up" period was included so that the choice of initial conditions did not impact upon the calibration period. The initial conditions adopted were that the soil moisture deficit within the loss module was assumed to be zero and the routing reservoirs were assumed to be empty. The behaviour of the calibrated model over an independent evaluation period forms a key component of identifying a "best" parameter vector within a catchment. The evaluation period was taken as the remainder of the period of record data available for a catchment. The mean length of record for the evaluation period within the catchment data set was 15 years.

### 6.4.2 The definition of the feasible parameter space and *a priori* determination of model parameters

The model parameter ranges that were used to define the feasible parameter space at the start of a calibration run are summarised within Table 6.1 for both model configurations. Of these parameters, the interception depth parameter, $\gamma$, for MODA, the split parameter, $\beta$, controlling the split of effective rainfall between quick and slow routing paths for MODB and catchment area were defined *a priori*. The remaining parameters were adjusted during calibration, as described within the previous section.

A basic assumption within both forms of the model developed for the regionalisation research is that of a closed catchment water balance, and thus an *a priori* estimate of catchment area is required. The estimation of catchment (or contributing) for a catchment was based on the topographic catchment boundary, the derivation of which is presented in

Chapter 5. The split parameter, β, for MODB was based on the value of the BFIHOST catchment characteristic (Chapter 5), as discussed within Chapter 4. The derivation of *a priori* estimates of the γ parameter for the interception model used within MODA was based upon the fraction extent of land cover classes within a catchment. The derivation of these estimates is presented within Appendix B.2.

The feasible parameter space for model parameters that were to be calibrated was defined on the following basis. The parameter space [25,1000] for Cmax was specified on the basis of the conceptual representation used within the model to simulate evapotranspiration. As discussed in Chapter 4, the formulation of the evaporation function is based on the concepts of the Penman drying curve theory and the concept of a root constant reflecting vegetation and soil type. The threshold at which the evaporation rate is impacted upon by a soil moisture deficit is approximately Smax/2. This threshold is roughly equivalent to a rooting depth. Grindley (1970) cites feasible rooting depth of between 25 and 250mm, depending on vegetation and soil type. As discussed in Chapter 4, Smax is a function of Cmax for both the Pareto distribution used in MODA (Smax=Cmax/(b+1)) and the uniform distribution used in MODB (Smax=Cmax/2). In terms of rooting depths, the Cmax feasible parameter space equates approximately to a maximum rooting depth parameter space of [13, 500] for MODA and [6, 250] for MODB, both of which encapsulate the Grindley values. In trails on various catchment types these limits were rarely approached by valid model shots during the first loop of the calibration scheme.

The feasible parameter space for the MODA drainage constant of proportionality, Kg, was more problematical to define, as in practice it varies widely without an obvious pattern. For this reason, a wide space was defined for the parameter and, in the first phase of the calibration scheme, the limits were treated as being fuzzy, as discussed in the previous section.

The initial, feasible parameter space for the time constant for the quick flow routing reservoir was defined through trials on a wide range of catchments, and represents a conservative feasible parameter space for this parameter within the UK. Again, these limits were treated as fuzzy limits. However, in practice, the initial limits were only revised in a small number of catchments.

The procedure for defining a catchment specific feasible parameter space for the time constant for the slow flow routing reservoir, Kb, was based on the analysis of recession periods within the observed flow records for a catchment. The full detail of this procedure is presented within Appendix B.3.

**Table 6.1 Feasible parameter ranges for MODA and MODB**

| Parameter (units) | Minimum | Maximum |
|---|---|---|
| $\gamma$ (MODA) (mm) | defined | *a priori* |
| Cmax (mm) | 25 | 800 |
| b MODA (none) | 0 | 4 |
| K1 (hr) | 2 | 200 |
| Kb (hr) | Catchment | Specific limits |
| Kg (MODA) (hr) | 1000 | 50000 |
| $\beta$ (MODB) (none) | defined | *a priori* |
| Area ($km^2$) | defined | *a priori* |

## 6.5 PARAMETER IDENTIFIABILITY FOR MODA AND THE DEVELOPMENT OF MODB

As discussed, the approach to calibration of the models for this study recognises that uniqueness of solution is unlikely to exist given the uncertainty associated with the input data and the sometimes-conflicting simulation requirements of different objective functions. The result is that there are ranges of stream flow simulations (with associated objective function values) that are regarded as equally likely realisations of the optimal simulation. As discussed, non-uniqueness of solution is a particular problem if parameters are to subsequently regionalised through relationships with catchment properties. The calibration scheme was used to assess the degree to which non-uniqueness exists among the parameters within the more complex MODA configuration of the model and to investigate relationships with catchment type. Evidence of structural identifiability problems led to the development of the simpler MODB model configuration by limiting the parameter identifiability problem at the expense of the conceptual structure and flexibility of the model. The results of this type of analysis are presented for four contrasting catchments within Appendix C.1 for MODA. Within this appendix, the four catchments are used to illustrate; the dependencies of the models on catchment type, the potential parameter covariance issues for MODA and the impact of errors in the input data upon the loss module parameters.

### 6.5.1 The case study catchments

The four case study catchment selected for the discussion represent the following climatological and hydrological permutations:

- dry- permeable:- The Bure at Ingworth (34003);
- dry-impermeable:- The Pinn at Uxbridge (39098);
- wet- permeable:- Sydling Water at Sydling St. Nicholas (44006);
- wet-impermeable:- The Gelyn at Cynefail (67010).

These particular catchments were selected as being representative of the quality of model fit obtained using MODA. The locations of the catchments are presented within Figure 6.4. Summary information about these catchments is presented in Table 6.2. The column entitled "% time SMD > 75mm" refers to the percentage of time that a simple modified Penman drying curve based model predicts a significant Soil Moisture Deficit (SMD) occurring, using the calibration period rainfall and PE input time series for the catchment. The drying curve model assumes that evaporation continues at the potential rate until the SMD exceeds a threshold rooting depth value, termed the Rooting Constant (RC). The evaporation rate then reduces gradually until it is zero at an SMD equivalent to twice the rooting depth. When the SMD is greater than the RC, the ratio of the actual evaporation rate (AE) to the potential evaporation rate (PE) is determined by

$$\frac{AE}{PE} = \left(1 - \frac{SMD - RC}{RC}\right) \qquad \textbf{(6.7)}$$

A rooting depth of 75 mm is equivalent to the reference rooting depth for short grass on good quality soils (Grindley, 1970). This Penman model is approximately equivalent to the evaporation model used in MODA for a catchment in which the catchment stores are a constant depth, and where the rooting constant is equal to Cmax/4 in the case of MODB.

**Table 6.2 Catchment summary information**

| Catchment | SAAR mm | Mean Flow $m^3s^{-1}$ | BFI | % time SMD > 75mm | Area $km^2$ | Comments |
|---|---|---|---|---|---|---|
| Bure at Ingworth (34003) | 686 | 1.12 | 0.83 | 36 | 164.7 | Two ogee profile weirs beneath bridge arches bypassed at 4.3 cumecs but maintains modularity. Rural land use catchment comprises sands gravels and loams. |
| Pinn at Uxbridge (39098) | 665 | 0.20 | 0.09 | 40 | 33.3 | Electromagnetic gauging station (overhead coil) in formalised trapezoidal section. Suburban catchment to the west of London, largely impermeable. Headwaters rise in countryside but substantial development down stream. |
| Sydling Water at Sydling St. Nicholas (44006) | 1030 | 0.18 | 0.86 | 18 | 12.4 | Crump profile weir. Modular under all flow conditions. Predominantly Lower Chalk with small outcrops of Middle and Upper Chalk forming the higher ground flanking the catchment. Mainly pastoral with some arable agriculture on flatter ground. |
| Gelyn at Cynefail (67010) | 2051 | 0.62 | 0.26 | 1 | 13.1 | Compound crump profile weir. Impermeable Lower Ordocician volcanics with occasional heavily indurated shales. Drift cover minimal. Upland pasture, rural. |

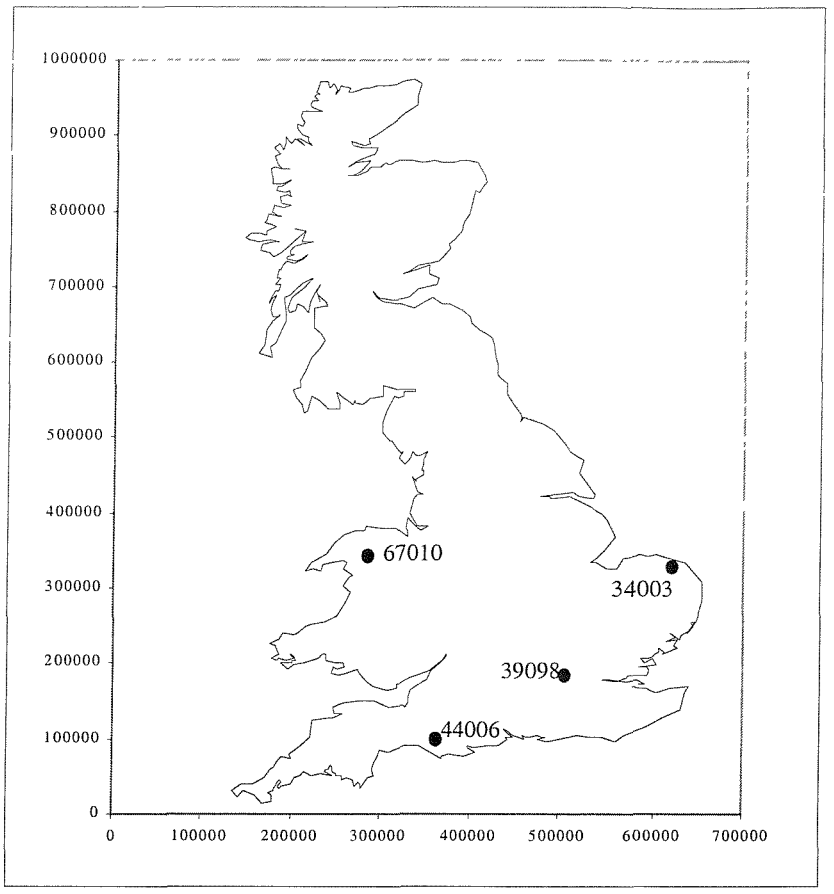



**Figure 6.4 The locations of the case study catchments.**

### 6.5.2 Summary of model behaviour for MODA

In all catchments there was very strong covariance exhibited between the loss module parameters Cmax and b and, to a lesser extent, between these parameters and Kg, the drainage time constant. A trade off between Cmax and b, in determining $\overline{C}$, and hence controlling evaporation losses was identifed. In the dryer catchments, Kg was also significantly co-variant with Cmax and b. There was also strong evidence of these parameters compensating for errors within the input data by adjusting the evaporation rate within the model. This is explained as follows, consider a fixed set of climatic input data that can potentially lead to the build up of evaporation limiting soil moisture deficits. A large value of $\overline{C}$ (high Cmax and low value of b) will minimise the time that evaporation limting soil moisture deficits occur, and hence will maximise the evaporation losses. Conversely, a low value of $\overline{C}$ will tend to minimise evaporation losses.

For a low value of Kg, the rate at which water is drained from the store will be high, although there is a negative feed back loop as the drainage rate is proportion to the depth of water held in storage with the soil store. This will have two effects; evaporation limiting soil moisutre deficits will build up more frequently and water will also be removed from the store, thus making it unavailable for evaporation. A low value of Kg will therefore tend to minimise evaporation losses. A high value of Kg will result in a reduction of drainage, and hence (for given values of Cmax and b) retention of water within the soil moisture store. This retion of water will lead to enhanced evaporation losses as signficant soil moisture deficits will not occur as frequently. This effect will be offset by enhanced direct runoff, and the fact that the drainage rate is proportion to the depth of water held in storage.

The role of the b and Kg parameters in controlling both the partitioning of effective rainfall between routing paths, and the relationship between precipitation and direct runoff (and hence the variation of EFF, LF_OBJ and BEQ95) was also identified. This role was more clearly demonstrated within high rainfall catchments. The role of these parameters within lower rainfall catchments was less clear, which is a consequence of the greater role that these parameters also play in controlling evaporation losses in this type of catchment.

The role of the routing reservoir time constants, Kb and K1, and hence the sensitivity of the model response to changes in these parameters also varied between catchments. For example, in base flow dominated catchments the model behaviour was sensitive to the value of Kb. Conversely, in impermeable catchments the model was relatively insensitive to the value of Kb - as a relatively small proportion of the effective rainfall was passing through the slow-flow reservoir. The parameter, Kb, is therefore only really identifiable if a significant proportion of the effective rainfall is passing through the slow-flow reservoir.

A related pattern was observed for the drainage time constant, Kg. In very impermeable catchments the absolute value of this parameter is unimportant, as long as it is large and thus preventing significant drainage into the slow-flow routing reservoir. This is not the case within permeable catchments. In permeable catchments the partitioning of effective rainfall between quick and slow flow paths is critically dependent on the value of Kg. To take an extreme case as an example, a Kg value of one would mean that the soil moisture

store would entirely drained at the beginning of each time step.

All of the formal objective functions BIAS, EFF and LF_OBJ are important in defining parameter values, and hence model fit for MODA. Furthermore, the BEQ95 statistic also appears to be of value in optimising the fit of the hydrograph at low flows. The BIAS and BEQ95 measures appear to be orthogonal, both to the other measures and to each other.

In all of the catchments there was evidence of a trade off between optimising EFF and LF_OBJ for MODA, indicating an inverse relationship between EFF and LF_OBJ. This behaviour is exacerbated through the choice of high and low BFI catchments for the discussion, but will probably still be present in catchments where there are both significant slow and fast routing paths. BEQ95 did not appear to be consistently related to any of the objective functions and appeared to be evaluating different aspects of model fit within different catchments.

### 6.5.3 Summary of model behaviour for MODB

The strong covariance between the parameters of the loss module within MODA is a significant issue for regionalisation. The soil moisture store parameters are not independent of one another, resulting in similar model fits for a wide range of parameter combinations. This is particularly important in catchments where significant SMDs are likely to build up. MODB was developed to reduce this parameter covariance within the soil moisture store parameters by including only one free parameter for calibration. This was achieved by assuming a uniform distribution of soil depths across the catchment and replacing the soil moisture dependent drainage term with a fixed partition of runoff from the soil moisture store, as discussed in Chapter 4. The coefficient controlling this split is analogous to the BFI. An *a priori* value for this coefficient was estimated for each catchment using the BFIHOST estimate (Chapter 5) derived from the fractional extent of HOST classes within each catchment. The interception store was omitted from MODB as the potential importance of the store in preventing the SMD being prematurely alleviated during the summer was negated by the omission of the soil moisture related drainage term.

This approach left three parameters for optimisation within the calibration scheme: Cmax, K1 and Kb. The trade off in reducing the number of free parameters in the loss module is

that the model is less flexible in closing a water balance. The model is also no longer dynamically routing effective rainfall between slow and fast routing paths on the basis of the status of the soil moisture deficit within the loss module. This is less conceptually attractive.

The number of catchments for which an acceptable water balance could be closed was reduced from 179 to 171. One of the catchment for which a water balance could not be closed was 44006. This is very likely to be either a consequence of errors in the input data or a violation of the closed water balance assumption, which can commonly occur within ground water fed catchments. This is discussed further within Appendix C. The behaviour of MODB is discussed with respect to the four case study catchments within Appendix C.2.

For MODB the parameters were uncorrelated with one another within all catchments. BIAS is strongly correlated with Cmax, as would be expected. There was no evidence of the trade off between the overall fit of the hydrograph and the fit at low flows for MODB, as was observed with MODA. This is a consequence of the use of a fixed split for partitioning effective rainfall. The time series fit statistics EFF, LF_OBJ and BEQ95 were strongly correlated with either Kb and/or K1. The degree of correlation was dependent on the catchment type. In impermeable catchments EFF was strongly correlated with the quick-flow reservoir, K1. In the permeable catchment EFF was correlated with Kb, as the majority of effective rainfall is routed through this store. The LF_OBJ function was correlated with both routing reservoir time constants. The degree of correlation was dependent upon the catchment type. In permeable systems the function and is associated with Kb and with both time constants in impermeable systems. There was evidence that the BEQ95 statistic can be correlated with all model parameters.

For MODB the BIAS objective function was strongly related to Cmax. The value of using all three EFF, LF_OBJ and BEQ95 to evaluate the fit of the simulated hydrograph was less certain. However the inter-correlations between these objective functions were not as high as the correlation between the individual functions and the appropriate time constant of the routing module. This indicates that they are evaluating the fit of different aspects of the simulated hydrograph.

## 6.6 SUMMARY, DEVELOPMENT OF THE EUCLIDEAN OBJECTIVE FUNCTION AND FINAL PARAMETER VECTOR SELECTION

The discussions within Appendix C and the summary of the results from the catchment studies presented within this chapter must be read in the context that all of the shots for a catchment were regarded as being acceptable. The focus of the analysis was to identify covariance, and hence identifiability issues within these acceptable shots. It was not to draw conclusions about the covariance relationships between parameters and objective functions over the whole of the feasible parameter space. An example of this is the case of the time constant for the routing reservoirs. Within the acceptable shots, the correlation between Kb and K1 and the objective functions LF_OBJ and EFF was small for some catchments. This is not the case over the whole of the feasible parameter space for these time constants. Calibration trials demonstrated that there is a strong correlation, particularly between LF_OBJ and Kb in permeable catchments and K1 and EFF within impermeable catchments. The lack of correlation within the range of acceptable shots is indicative of the lack of sensitivity of the model to smaller scale variations in the time constants compared with equivalent variations in the other parameters. The sensitivity of the model behaviour to individual parameter values, and hence the identifiability of the parameters, was clearly linked to the catchment type and climatic regime. This is also true with regard to parameter covariance issues. For example, the loss module parameters for MODA were much more covariant in catchments in which the parameters control the evaporation losses from the soil moisture store.

The calibration scheme ensures that model parameter vectors can only be drawn from regions of the feasible parameter space in which the model can close a water balance. If the errors in input data are appreciable, leading to potential over-estimation or under-estimation of mean flow, the model will tend to compensate for this by selecting loss module parameters that will simulate high or low evaporation losses. As these loss module parameters also control the generation of direct runoff, and in the case of MODA the drainage from the soil moisture store this can result in a poor simulation of the time dependency of stream flows.

The experience with MODA and MODB does imply that if BIAS is to be used as an objective function then, if more than one parameter can influence the simulation of evaporative losses there are likely to be covariance issues between these parameters in low rainfall catchments.

As discussed in Section 6.2, each of the 300 model parameter vectors can be regarded as, potentially, being the optimal, or near optimal solution for the model - given the uncertainties associated with the input data and conflicting simulation requirements of objective functions. For the regionalisation studies it is essential to represent each catchment by one parameter vector. The predictive capacity of the model, when applied to an independent evaluation period of record is a good mechanism for identifying a stable model parameter vector. For the purposes of this research, the identification of a stable, near optimal model parameter vector is more important than the identification of an optimal model parameter vector for a specific period of record - which maybe an artefact of the input data for that period.

The approach adopted to derive this vector for each catchment was to select a best compromise parameter vector through the use of a Euclidean based objective function. This function is based on the sum of distances of the value of each of the BIAS, EFF, LF_OBJ and BIAS statistics within the calibration period and the departure from these over the evaluation period. The Euclidean objective function for a model vector was derived using

$$EOF = \frac{N_C COF + N_E EVOF}{N_C + N_E}, \tag{6.8}$$

where COF is the calibration period function given by

$$COF_i^2 = \left( \frac{|bias|_i - |bias|_{min}}{|bias|_{max} - |bias|_{min}} \right)^2 + \left( \frac{OBJ\_LF_i - OBJ\_LF_{min}}{OBJ\_LF_{max} - OBJ\_LF_{min}} \right)^2 + \left( \frac{(1-EFF)_i - (1-EFF)_{min}}{(1-EFF)_{max} - (1-EFF)_{min}} \right)^2 + \left( \frac{|BEQ95|_i - |BEQ95|_{min}}{|BEQ95|_{max} - |BEQ95|_{min}} \right)^2, \tag{6.9}$$

and EVOF is the evaluation period function given by

$$EVOF_i^2 = \left(\frac{bias_i - bias_{min}}{bias_{max} - bias_{min}}\right)^2 + \left(\frac{OBJ\_LF_i - OBJ\_LF_{min}}{OBJ\_LF_{max} - OBJ\_LF_{min}}\right)^2 + \left(\frac{(1-EFF)_i - (1-EFF)_{min}}{(1-EFF)_{max} - (1-EFF)_{min}}\right)^2 + \left(\frac{BEQ95_i - BEQ95_{min}}{BEQ95_{max} - BEQ95_{min}}\right)^2 . \quad \textbf{(6.10)}$$

Within these functions, the values of the original objective functions for all valid parameter vectors are normalised to vary in the range [0,1] within each period. A value of zero represents a perfect simulation for the function concerned. The dashes for the fit statistics within the EVOF function denote that the value is the difference between the value of the objective function during the calibration period and the evaluation period. The BIAS, EFF and BEQ95 statistics were normalised prior to subtracting the values for the calibration period from those for the evaluation period. The variables $N_C$ and $N_E$ are respectively the number of years in the calibration period and evaluation period respectively. This approach is directed at identifying the best compromise between the fit of the model during the calibration period and the stability of the model fit over the evaluation period. The weight given to COF and EVOF is based on the respective record lengths of the periods. The parameter values and objective function values for the shots selected in the case study catchments using the EOF are presented in Table 6.3 for MODA and MODB. Values are presented for both the calibration and evaluation periods.

These results demonstrate that the selected vectors produce relatively stable estimates over both the calibration and evaluation periods. Noting that for two of the catchments the evaluation period is restricted to two years of record, the models conserve mass over both the calibration and evaluation periods. In all cases, the fit of the model at low flows (LF_OBJ and BEQ5) is better over the evaluation period than the calibration period, a consequence of the calibration period containing the drought years of the early 1990s. The EFF values do not follow a consistent pattern between the calibration and evaluation periods.

The performance of the calibrated parameter vectors selected using this approach for these catchments is discussed in more detail within Chapter 8, Section 8.1. In this section the performance of regionalised parameter estimates for these catchments is compared to the performance of the calibrated model parameters. The discussion also includes a visual assessment of simulated hydrographs.

**Table 6.3 Selected parameters and corresponding objective function values for MODA and MODB**

| | Parameter Values | | | | | Objective function values | | | | | | | | years of record | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Calibration | | | | Evaluation | | | | | |
| | Cmax | B | K1 | Kb | Kg | LF_OBJ | BIAS | EFF | BEQ95 | LF_OBJ | BIAS | EFF | BEQ95 | NC | NV |
| **MODA** | | | | | | | | | | | | | | | |
| **34003** | 793 | 1.07 | 57 | 1639 | 6374 | 27 | 2 | 0.526 | -20 | 20 | -3 | 0.54 | 6 | 10 | 10 |
| **39098** | 373 | 0.95 | 26 | 144 | 51601 | 63 | 1 | 0.684 | -11 | 53 | -7 | 0.43 | -6 | 10 | 2 |
| **44006** | 433 | 0.03 | 384 | 505 | 7365 | 25 | 0 | 0.60 | 24 | 24 | 3 | 0.72 | 28 | 10 | 10 |
| **67010** | 110 | 3.04 | 30 | 496 | 762 | 46 | -3 | 0.734 | -6 | 45 | -2 | 0.76 | 4 | 10 | 2 |
| **MODB** | | | | | **Beta** | | | | | | | | | | |
| **34003** | 251 | 1 | 112 | 1770 | 0.779 | 35 | 0 | 0.174 | -47 | 30 | 0 | 0.29 | -22 | | |
| **39098** | 386 | 1 | 32 | 398 | 0.174 | 75 | 0 | 0.733 | -22 | 67 | -4 | 0.68 | -5 | | |
| **67010** | 37 | 1 | 22 | 492 | 0.25 | 48 | 2 | 0.78 | -9 | 51 | 3 | 0.81 | 9 | | |

# 7 The development of regionalised model parameters

## 7.1 INTRODUCTION

This chapter presents the development of predictive models that link the rainfall runoff model parameters to catchment characteristics, thus facilitating the application of the rainfall runoff models to ungauged catchments. As discussed in Chapter 1, the process of identifying models that describe the relationship between a variable and co-variant catchment characteristics is commonly called regionalisation. The evaluation of the utility of the models for predicting rainfall runoff model parameters that yield acceptable stream flow simulations is presented within Chapter 8.

Catchment characteristics/descriptors are defined as the physical characteristics of a basin that control the generation of stream flow, corresponding to precipitation inputs and evaporative losses from the catchment surface, and climatic characteristics that can be used to summarise the spatial patterns within the climate data over the catchment. The catchment characteristics/descriptors used in this research are fully described within Chapter 5. Conceptually correct models for predicting model parameters were sought in the regionalisation studies; that is, models that can, conceptually, describe cause and effect.

Two approaches were developed and evaluated in detail, these were a multivariate regression approach and a Region of Influence approach (ROI). In the regression based approach, the parameters within a rainfall runoff model parameter vector are assumed to be independent of one another, this is contrary to the evidence from the catchment specific studies discussed in Chapter 6. As discussed in Chapter 1, multivariate regression analysis has been used by other researchers seeking to relate model parameters to catchment characteristics, and has also been widely used to develop simple, statistically based hydrological models for predicting natural and artificially influenced flow statistics. Linear regression is a special case of the family of Generalised Linear Models (GLM) (Cox and Hinkley, 1974). More sophisticated models can be used from this family by ascribing a probability distribution to the dependent variable, in this case a model parameter. Some function, $f$, known as the link function of the distribution parameters, is then modelled as a linear combination of the independent predictor variables. The use of GLM was rejected

because of the demonstrable covariance between model parameters within individual catchments. The degree of covariance varies as an undefined function of catchment type, as demonstrated empirically within Chapter 6. It was not therefore possible to define a defendable underlying distribution for the individual model parameters that were to be regionalised.

The parameters within individual vectors are not assumed to be independent in the ROI approach. In this approach, likely parameter vectors are selected for the target-ungauged catchment from vectors for a pool of source catchments. The selection is based upon the catchment characteristic similarity of the source catchments to the target catchment. The ROI approach is a logical extension of cluster analysis. The objective of cluster analysis is to place elements in to self-similar groups to make some value judgements about the properties of the self-similar groups. In this context, similarity would be assessed on the basis of catchment characteristics that are co-variant with the model parameters. The objective in the ROI approach was to develop a method for identify a group of catchments with similar characteristics, and hence (ideally) model parameters, with the catchment of interest located at the centre of the group.

The data sets used for this analysis consisted of the calibrated MODA and MODB parameter vectors and catchment characteristics for the catchments for which the calibration scheme (discussed in Chapter 6) was successfully applied. Section 7.2 of this chapter presents a discussion of the calibrated model fits obtained for both model structures across the catchment data sets. A discussion of the variation in catchment characteristics/descriptors for these catchments, and the relationships between individual characteristics/descriptors is presented within Section7.3. The development of multivariate regression based models for predicting model parameters is presented in Section 7.4 for both models. The section presents both the methodological approaches used within the derivation of these parameter estimation models and the models derived. In practice, the stream flow simulations obtained using the ROI approach were not as good as the simulations obtained using the multivariate regression models. The derivation of the ROI analysis and the application to the estimation of parameters for MODA is therefore presented within Appendix D. The results from the studies presented within this chapter are summarised in Section 7.5.

## 7.2 AN EVALUATION OF THE CALIBRATED MODEL FITS ACROSS THE STUDY CATCHMENTS

As discussed in Chapter 6, it was possible to obtain calibrated model fits that closed an effective water balance in 179 catchments using MODA and 170 catchments using MODB. Histograms of the values of the BIAS and EFF objective functions and the $R^2$ and BEQ95 goodness of fit statistics obtained within these catchments, for both model configurations are presented in Figure 7.1 for the calibration period and Figure 7.2 for the evaluation periods. These statistics are fully described within Chapter 6. The LF_OBJ objective function used in the calibration period is not normalised, and is therefore not easy to interpret in the context of inter-catchment comparisons. The fit at low flows is represented by the BEQ95 statistic only. Catchments in which the BEQ95 error was greater than 300% were omitted from the histograms for plotting purposes. For MODA there were nine catchments with BEQ95 values greater than 300%, and for MODB there were three catchments. The large BEQ95 errors observed within these catchments are discussed further within Section 8.2 of Chapter 8.

The $R^2$ statistic, commonly used to describe the fit of a model, is presented here as a comparison to the EFF statistic. Both histogram figures demonstrate that higher values are observed for $R^2$ than for EFF, which is indicative of the complex error structures within stream flow data, as discussed in Chapter 6. Summary statistics (median and 68% confidence interval (c.i.) limits) for the BIAS, EFF and BEQ95 statistics are presented in Table 7.1 for both the calibration and evaluation periods. Summary statistics about the percentage departures, or differences, between the fit in the calibration period and evaluation period for these statistics are also presented within the table.

To differentiate between statistics drawn from different periods, the statistic name is suffixed with underscore "c" to represent a calibration period statistic and "v" to represent a verification, or evaluation period statistic. The suffix "d" is used to represent the percentage differences in the value of a statistic between the two periods.

Spearman rank correlation analysis was used to demonstrate that the objective functions used were uncorrelated across the catchments. The same analysis was used to demonstrate than there was no consistent dependency of the performance of the model (as measured by

LF_OBJ, BIAS, EFF and BEQ95) on the permeability of the catchments (as represented by BFI) or how wet the catchments are (as represented by SAAR). The fits for individual statistics are discussed within the following sub-sections.

**Table 7.1 Summary statistics for model fits over the calibration and evaluation periods**

| | | MODA | | | MODB | | |
|---|---|---|---|---|---|---|---|
| | | | 68% c.i | | | 68% c.i | |
| | | Median | u.lim | l.lim | median | u.lim | l.lim |
| **Calibration period** | BIAS_c | 0 | 1 | -2 | 1 | 3 | -1 |
| | $R^2$_c | 0.86 | 0.91 | 0.78 | 0.86 | 0.90 | 0.79 |
| | EFF_c | 0.73 | 0.81 | 0.57 | 0.71 | 0.80 | 0.59 |
| | BEQ95_c | 16 | 78 | -7 | -6 | 35 | -28 |
| **Verification period** | BIAS_v | -6 | 0 | -11 | -4 | 2 | -10 |
| | $R^2$_v | 0.83 | 0.88 | 0.75 | 0.84 | 0.89 | 0.76 |
| | EFF_v | 0.67 | 0.77 | 0.55 | 0.68 | 0.77 | 0.54 |
| | BEQ95_v | 22 | 75 | -4 | -6 | 27 | -28 |
| **percentage differences** | BIAS_d | -4 | -2 | -10 | -3 | 0 | -9 |
| | EFF_d | -4 | 1 | -8 | -3 | 1 | -7 |
| | BEQ95_d | -2 | 26 | -27 | 1 | 24 | -21 |

Note: negative percentage differences indicate the calibration fit for the statistic is better

### 7.2.1 The BIAS statistic

The calibration procedure constrained BIAS over the calibration period to within ±3% for the majority of catchments. Within this range, MODA has a tendency to under-estimate MF (negative BIAS) over the calibration period and MODB has a tendency to over-estimate MF (positive BIAS). This is confirmed by the 68% c.i for BIAS for each model. Both models have a tendency to under-predict mean flow over the evaluation period, this is possibly a function of the calibration period (generally the 10-year period of record preceding 1997) being dryer than the evaluation period (generally 1976-1985) for most catchments. MODB tends to under predict mean flow slightly less than MODA.

### 7.2.2 The EFF statistic

The performance of both models with respect to EFF is similar over the calibration period. At 10 years, the calibration period is longer than for conventional model applications in which a 2 to 3 year period is more commonly used. The calibrations of MODA within the case study catchments, used in Chapter 6 to explore the behaviour of the models, were used to explore the impact of record length on the calculation of EFF. The 10-year calibration periods were partitioned into two-year blocks and a value of EFF was calculated for each block. The mean of these two year EFF values, and the EFF values calculated over the entire calibration period are presented for each catchment within Table 7.2. This demonstrates that, in all catchment the mean EFF values calculated from the two-year blocks are much higher than the corresponding 10-year value, despite being calculated using exactly the same data. This is a direct consequence of the complex error structures introduced by the strong serial correlation within both the observed and simulated time series.

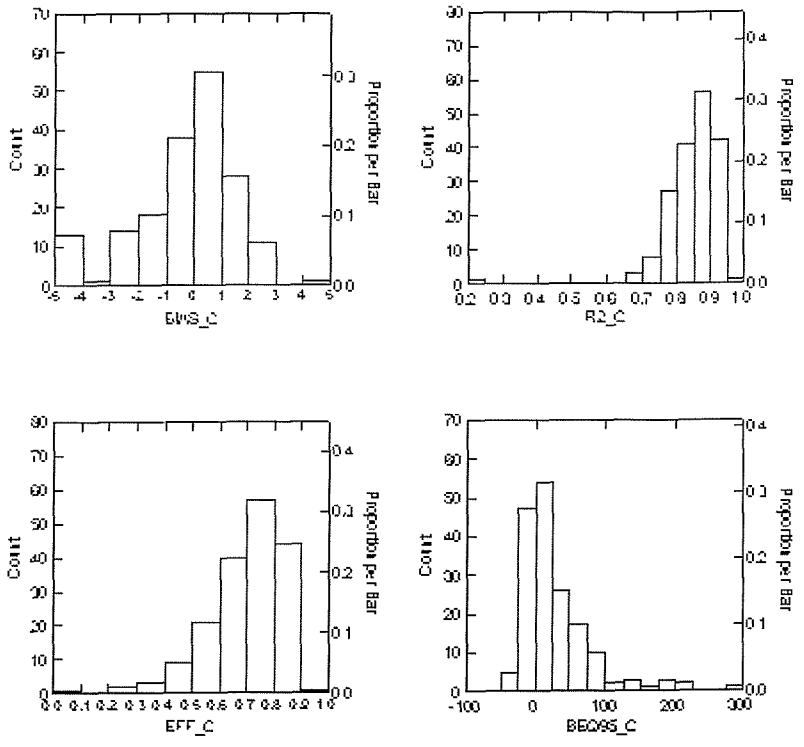
Returning to the values of EFF obtained for the calibration and evaluation periods over all the catchments used in the study, the performance of both models is good with 84% of catchments having EFF values over the calibration period of greater than 0.57 for MODA and 0.59 for MODB. Six catchments have low values of EFF (<0.4) for MODA. The station numbers for these catchments are 28008, 42014, 34002, 55028 and 28008. The EFF value for these catchments all improved to above 0.4 over the evaluation period, although the EFF values for a number of other catchment dropped below 0.4 during the same period. Four catchments have poor values of EFF (39029, 31017, 34003, 42023) for MODB, again the EFF values over the evaluation period improved. Looking at the distribution of percentage changes over all catchments, the degradation of model fit over the evaluation period is small with 68% of catchments having EFF values that are within 8% of the calibration period for MODA and 7% for MODB. The upper limit for the 68%c.i. is positive for both models at 1%. This indicates that, in at least 16% of catchments the value of EFF was higher over the evaluation period than over the calibration period. EFF values for MODA are slightly higher over the calibration period than for MODB. Over the evaluation period the MODB model fits are slightly more stable than the MODA fits with respect to EFF.

**Table 7.2 Impact of record length on the value of EFF**

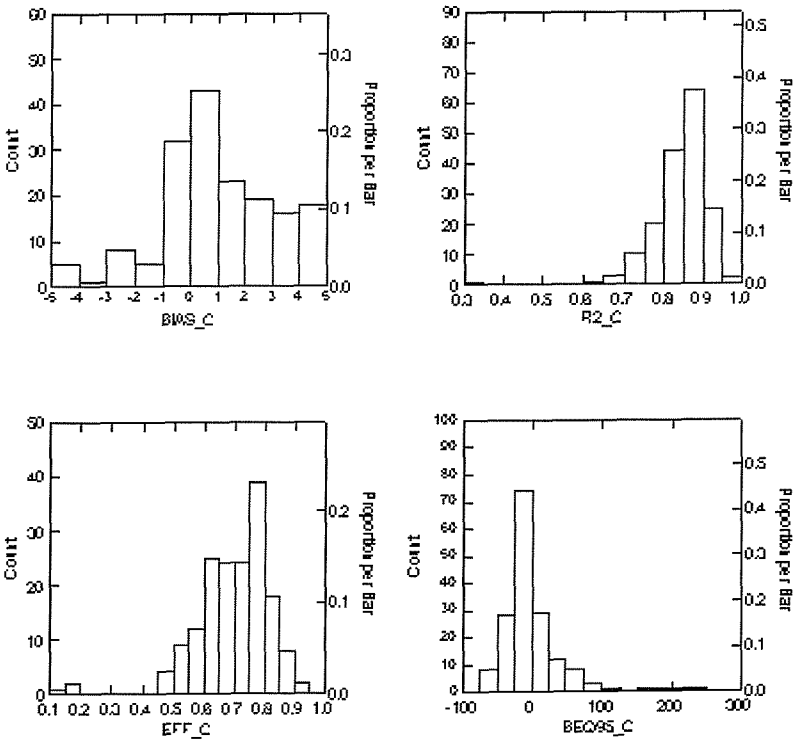
| Catchment | EFF full calibration period | Mean EFF from two year blocks |
|---|---|---|
| 34003 | 0.53 | 0.91 |
| 39098 | 0.68 | 0.82 |
| 44006 | 0.60 | 0.89 |
| 67010 | 0.73 | 0.78 |

### 7.2.3 The BEQ95 Statistic

The histograms and 68% c.i for BE95 demonstrate that MODA tends to over-estimate Q95 flows over both the calibration and evaluation periods. The median value of BEQ95 is 16% for MODA, and the 68% upper limit is high at 78%. The performance of MODB is better (smaller 68%c.i. and less biased), with a tendency to under-predict slightly. Nine catchments with errors greater than 300% over the calibration period have been excluded from the histogram for MODA, and 3 catchments excluded from the MODB histogram. Only two of the MODA exclusions (39065: 338% and 47005: 310%) were amongst the nine catchments that were excluded from the calibration analysis for MODB, due to an inability to close a water balance. All of the excluded catchments had low Q95 values. In these catchments a given absolute BEQ95 error will represent a larger percentage difference. Analysis of the observed data for the catchments with the three largest errors (36009, 40027, and 39042) showed that all three have a tendency to dry up during the summer months. The median and 68%ci for percentage differences between the calibration and evaluation periods demonstrate that the performance of both models in modelling Q95, and hence low flows is better over the evaluation period than the calibration period. This is a consequence of the low flows being generally higher over the wetter evaluation period.

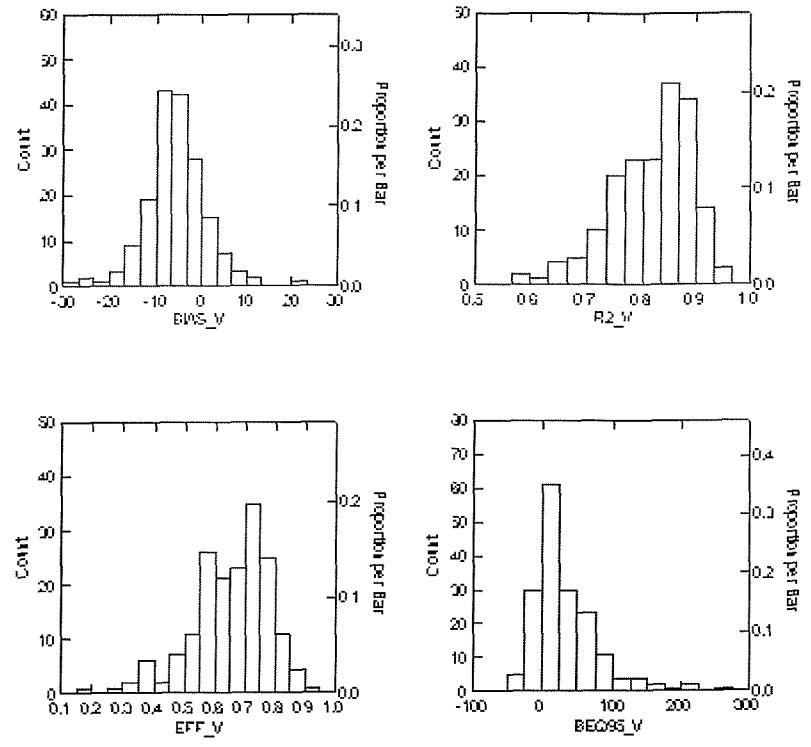



**(a) MODA**

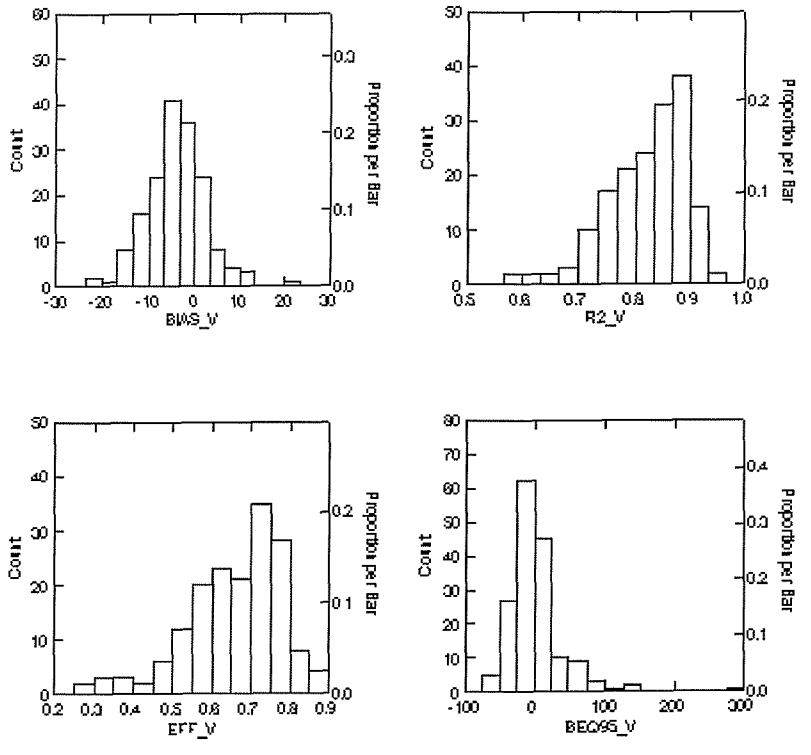



**(b) MODB**

**Figure 7.1 Objective function histograms for the calibration period.**




**(a) MODA**




**(b) MODB**

**Figure 7.2 Objective function histograms for the evaluation period.**

### 7.2.4 Summary

The performance of both model configurations is fairly similar across the study catchments. The majority of model fits are very good; particularly when the long (10 year) calibration period used is considered. MODA is better at predicting mean flow and MODB is better at predicting low flows. Both models are stable when applied to an evaluation period; the mean length of record for the evaluation period was 15 years. A factor in this result is that the evaluation period was generally wetter than the calibration period for the majority of catchments, thus making the modelling of soil moisture behaviour less critical. The fits for MODB are generally more stable than those for obtained for MODA. This indicates that MODA may be over specified. In some catchments, poor fits for either the EFF or BEQ95 statistics were obtained over the calibration period. As the fits improved over the evaluation period, these catchments were not rejected from the regionalisation studies in the interests of maintaining sample size for the analysis.

The use of Spearman rank correlation analysis demonstrated that there was no significant inter-correlation between the objective functions used across the catchments. Furthermore and there were no consistent dependencies of model performance on the permeability and wetness of the catchments.

## 7.3 THE CHARACTERISTICS OF THE CASE STUDY CATCHMENTS

The catchment characteristics/descriptors used for the study can be broadly classed into three types:

- Topographical;
- Climatological;
- Soils and Land cover.

The catchment characteristics/descriptors, their derivation and nomenclature are described in more detail within Chapter 5. The characteristics within the topographic and climatic classes are continuous variables. The HOST (soils) and Land Cover characteristic are fractional extents of discrete classes (30 and 25 respectively), and are constrained so that fractional extents within a catchment sum to one. This places a constraint on regression

modelling using these characteristics, as the fractional extents of classes within a catchment are not independent of one another. This issue is discussed further in Section 7.4.

Spearman rank correlation analysis was used to identify the significant correlation coefficients between the continuous catchment characteristics/descriptors. These are presented in Table 7.3 for the MODA catchments. The correlation values are grouped into bands, based on values and highlighted in colour, for ease of interpretation. A scatter plot matrix for the characteristics is presented in Figure 7.3.

**Table 7.3 Spearman rank correlation between catchment characteristics for MODA**

| | SAAR | SAAPE | PP | SPR | BFI | AREA | LDP | DPLBAR | DPLCV | ALTBAR | DPSBAR | ASPBAR | ASPVAR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SAAR** | 1.00 | | | | | | | | | | | | |
| **SAAPE** | -0.58 | 1.00 | | | | | | | | | | | |
| **PP** | -0.90 | 0.79 | 1.00 | | | | | | | | | | |
| **Q5** | 0.18 | -0.03 | -0.09 | | | | | | | | | | |
| **Q95** | -0.29 | 0.11 | 0.22 | | | | | | | | | | |
| **SPR** | 0.43 | -0.50 | -0.53 | 1.00 | | | | | | | | | |
| **BFI** | -0.43 | 0.34 | 0.43 | -0.77 | 1.00 | | | | | | | | |
| **RO** | 0.97 | -0.67 | -0.93 | 0.48 | -0.43 | | | | | | | | |
| **AREA** | -0.04 | -0.29 | -0.13 | -0.10 | 0.22 | 1.00 | | | | | | | |
| **LDP** | -0.02 | -0.31 | -0.17 | -0.08 | 0.18 | 0.96 | 1.00 | | | | | | |
| **DPLBAR** | -0.02 | -0.30 | -0.16 | -0.08 | 0.19 | 0.96 | 0.99 | 1.00 | | | | | |
| **DPLCV** | 0.19 | -0.18 | -0.20 | 0.23 | -0.16 | -0.03 | 0.06 | -0.02 | 1.00 | | | | |
| **ALTBAR** | 0.73 | -0.75 | -0.84 | 0.47 | -0.28 | 0.17 | 0.20 | 0.19 | 0.24 | 1.00 | | | |
| **DPSBAR** | 0.75 | -0.61 | -0.76 | 0.30 | -0.15 | 0.05 | 0.05 | 0.05 | 0.16 | 0.83 | 1.00 | | |
| **ASPBAR** | -0.07 | 0.20 | 0.11 | -0.07 | 0.07 | -0.19 | -0.17 | -0.17 | 0.00 | -0.16 | -0.09 | 1.00 | |
| **ASPVAR** | -0.05 | 0.19 | 0.11 | 0.04 | -0.15 | -0.61 | -0.54 | -0.54 | 0.02 | -0.11 | -0.15 | 0.02 | 1.00 |

$0.5 \le r^2 < 0.6$ $0.6 \le r^2 < 0.8$ $r^2 \ge 0.8$

The histograms from the scatter plot matrix demonstrate that the majority of the characteristics are approximately normally distributed. The exceptions are the potential for soil moisture deficits (PP) and catchment area (AREA). The histogram for AREA demonstrates the catchment data set is skewed towards small catchments. The histogram for PP demonstrates that a significant number of the catchments are high rainfall catchments that have a zero PP value. Some of the relationships between characteristics are fairly linear, for example BFIHOST and SPRHOST, others are either non-linear, or indicate that values for one characteristic are constrained with respect to the variation with

a second characteristic. For example, high values of the variability in catchment (ASPVAR) do not occur for catchments with long drainage paths (LDP), this demonstrates that the variability in aspect is always low in the larger catchments. The correlation between SPRHOST and BFIHOST led to the use of the HOSTRES characteristic, described in Chapter 5. HOSTRES is uncorrelated with both SPRHOST and BFIHOST.

Principal Components Analysis (PCA), also known as empirical orthogonal function analysis, was used to explore how the continuous characteristics could be summarised to explain the variance within the data set. PCA is used to construct linear combinations $Z_1$, $Z_2$, $Z_3 \ldots Z_p$ of $p$ variables $X_1$, $X_2$, $X_3 \ldots X_p$ (in this 13 characteristics (including HOSTRES)) that are uncorrelated where

$$Z_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + a_{14}X_4 + .... + a_{1p}X_p$$

$$a_{11}^2 + a_{12}^2 + a_{13}^2 + a_{14}^2 + ..... + a_{1p}^2 = 1 \qquad \textbf{(7.1)}$$

As they are uncorrelated the combinations are measuring different dimensions within the data. The coefficients, $a_{ip}$, are a measure of the contribution of variable, $X_p$, to $Z_i$. A PCA involves finding the Eigen values of the sample covariance matrix for $X_p$. For the PCA the normal practice of coding $X_p$ so as they have means of zero and variance of one was adopted. This avoids one, or more variables having an undue influence on the principal components as a result of scale effects.

Where two or more variables within a component analysis are correlated with one another, the coefficients, a, are not independent and are therefore not easily interpretable. To resolve this, it is common practice to look at the loading for component variables, $X_i$, with the principal component $Z_i$. The loading for a component variable is the correlation of that variable with the principal component. Plotting the component loadings may reveal that a variable has a sizeable loading for more than one principal component. One refinement is to use rotated loadings with the PCA (termed orthogonal PCA). In this analysis, if a variable has a sizeable loading for more than one PC, the plot axes for the PC are rotated with the aim of maximising the component loading for a variable with one PC and minimising it with respect to the others. This generally maximises the larger un-rotated loadings for a PC and minimises the smaller loadings.

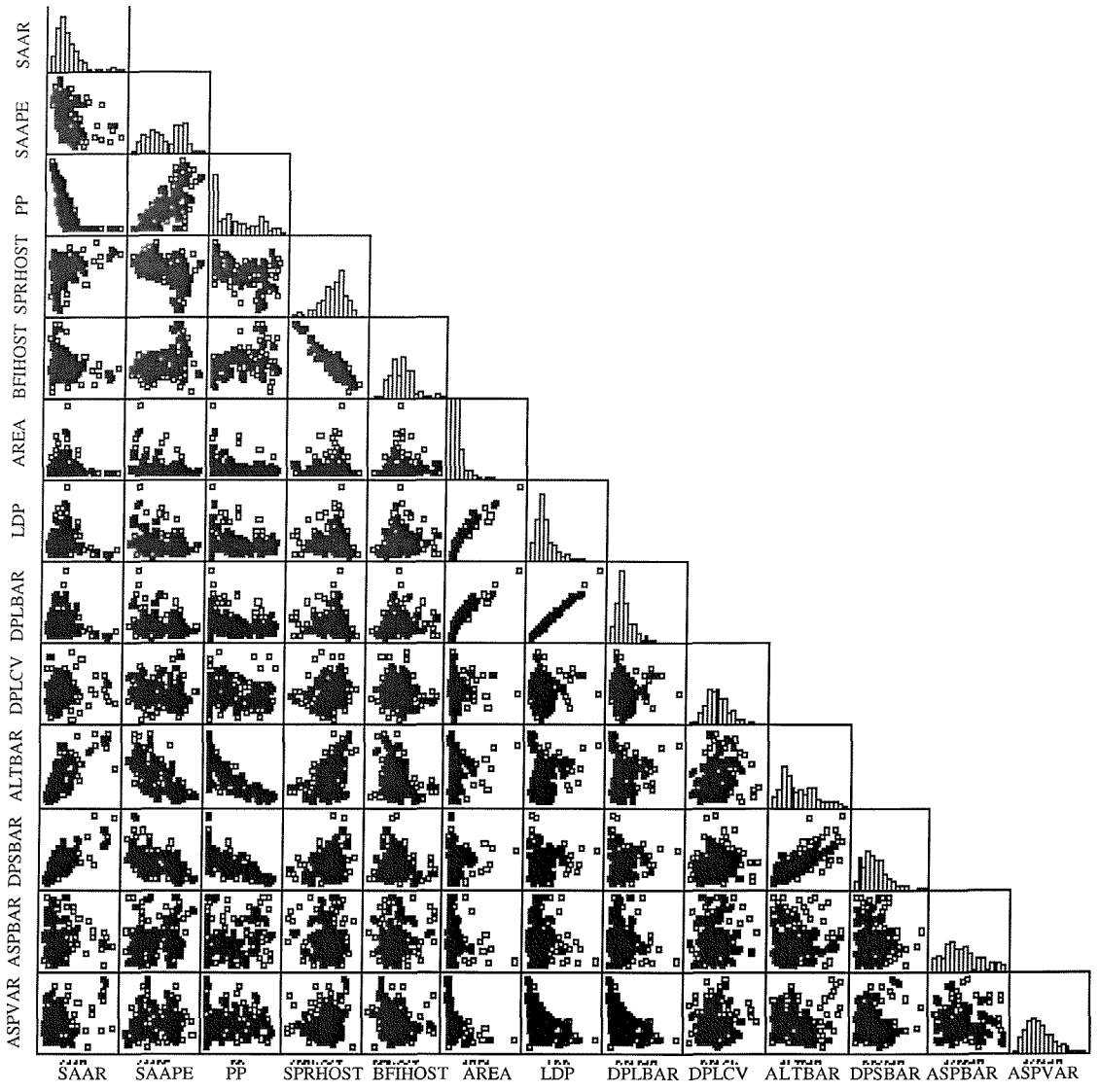



**Figure 7.3 Scatter plot matrix for MODA catchment characteristics.**

The sum of Eigen values for all PC is equal to the sum of the trace (diagonal) of the covariance matrix. As the diagonal of a covariance matrix contains the variance of each of the variables within the sample, the Eigen value for a PC is the variance of the PC. The scree plot, presented in Figure 7.4 (Eigen values plotted as a function of PC number) for the PCA of catchment characteristics, demonstrates that there are three PC with Eigen values greater than one. The rotated component loadings for these PC, together with the Eigen values, $\lambda$, and the percentage of the total variance explained by the components are summarised in Table 7.4. The variables with rotated component loadings greater than 0.6 are highlighted in bold. The three PC explain 69% of the total variance within the catchment characteristic/descriptor data set.

The characteristics associated with climate (SAAR, SAAPE, and PP), catchment altitude (ALTBAR) and mean slope (DPSBAR) are dominant within the first component. These variables are also strongly correlated with one another. It is apparent from the signs of the loading that the first component is large for wet, high, steep catchments with lower potential for evaporation (related to altitude) and negligible potential for developing significant soil moisture deficits. This component explains nearly 30% of the total variance within the data set. The second component is primarily associated with variables that are related to catchment size and configuration. This component is larger for smaller (AREA, LDP) catchments that have a more variable aspect (ASPVAR) and that tend not to be linear in shape (DPLBAR). The third component is associated with BFIHOST, SPRHOST and HOSTRES (note HOSTRES is uncorrelated with either of the other two characteristics). This component is larger for impermeable catchments that tend to have a higher standard percentage runoff than expected. This last component only explains 16% of the variance within the sample data set.

The dominance of the climatological, altitude and slope characteristics in explaining the variance within the data set is a reflection of both the strong correlation relationships between these variables, and the large variation in the magnitude of these variables across the UK. This variation may be in excess of one order of magnitude. This large variation is also seen within the LDP and AREA characteristics within the second component. The relative importance of the PC in explaining the variation within the catchment data set does not imply that the characteristics within the components will carry the same importance in controlling the variability in model parameters. However, the fact that ASPBAR and DPLCV are not useful in explaining the variation within the catchment characteristics data set does imply that these are unlikely to be useful co-variates for the rainfall runoff model parameters. The PCA analysis does provide useful guidance on determining whether it is justifiable to include more than one characteristic from a particular PC in the regression analysis between model parameters and characteristics.

**Table 7.4 PCA analysis results**

| | Principal Components | | |
|---|---|---|---|
| | **1** | **2** | **3** |
| DPSBAR | **0.90** | 0.00 | -0.04 |
| ALTBAR | **0.88** | -0.10 | 0.22 |
| PP | **-0.84** | 0.16 | -0.28 |
| SAAR | **0.84** | 0.23 | 0.14 |
| SAAPE | **-0.74** | 0.30 | -0.27 |
| LDP | 0.05 | **-0.98** | 0.06 |
| DPLBAR | 0.04 | **-0.98** | 0.04 |
| AREA | 0.10 | **-0.93** | 0.06 |
| ASPVAR | 0.04 | **0.60** | 0.22 |
| BFIHOST | -0.23 | -0.14 | **-0.92** |
| SPR | 0.42 | 0.10 | **0.77** |
| HOSTRES | 0.43 | -0.09 | **-0.61** |
| DPLCV | 0.14 | -0.09 | 0.28 |
| ASPBAR | -0.07 | 0.21 | -0.20 |
| Eigen values | 4.00 | 3.41 | 2.20 |
| % tot Variance | 29 | 24 | 16 |

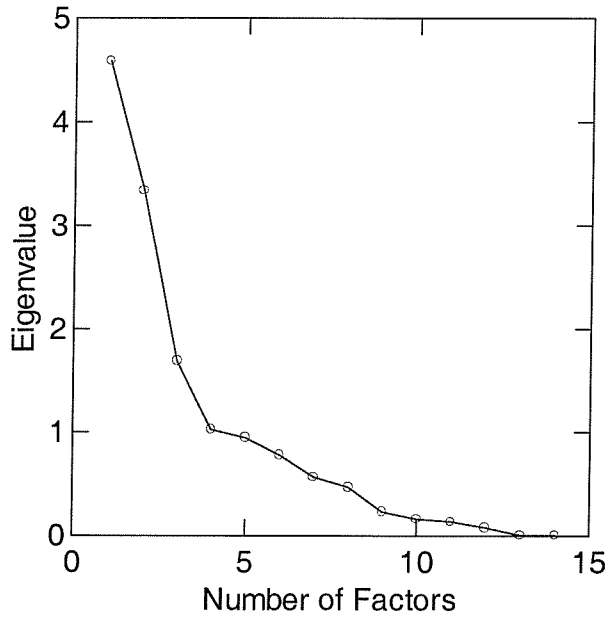



**Figure 7.4 Scree Plot for catchment characteristic PCA.**

## 7.4 THE DEVELOPMENT OF LINEAR REGRESSION BASED MODELS FOR PREDICTING MODEL PARAMETERS

### 7.4.1 Methodology

The primary methodology used for developing regression relationships between model parameters and catchment characteristics was to seek relationships between the continuous characteristics and the parameters before seeking relationships between the parameters and the discrete characteristics. If the HOST derived characteristics (BFIHOST or SPRHOST) proved to be dominant in a particular relationship, then models relating the parameter directly to the fractional extents of HOST were developed to maximise the relationship.

As an example, scatter plot matrices for the relationships between the continuous catchment characteristics and model parameters for MODA are presented in Figure 7.5. These demonstrate that, for some characteristics there are no obvious relationships with model parameters, for some the relationships are linear and for some the relationships appear to be non-linear. Additionally, the distributions for some characteristics, and some parameters appear to be skewed with evidence of potential outliers. Similar results were observed for relationships between catchment descriptors and model parameters. In developing the regression models, logarithmic or square root transformations were applied to the dependent and/or the independent variables, as appropriate, to linearise relationships, where necessary and to transform distributions to a better approximation of uniformity and thus reducing heteroscedacity issues within the regression analysis. The value of transforming variables was assessed through graphical and Pearson correlation analysis prior to the regression analysis.

In the case of fitting regression relationships between parameters and the fractional extent of HOST classes, the regressions were constrained to have a zero intercept, as the fractional extents must sum to unity. This was essential to resolve the non-independence of fractional extents, as represented by high condition indices within the analysis. By extension, other characteristics could not be included directly within fractional extent based regressions. To resolve this, the relationship between the parameter in question and fractional extent of HOST classes was maximised in a first stage. A second model was then fitted to the

regression residuals using the continuous characteristics. In all instances where this was attempted the second stage failed to yield useful (significant) relationships. The use of fractional extents of LANDCOVER and a simple combination of LANDCOVER and HOST were also investigated as an alternative to HOST for the loss model parameters. These parameters conceptually, could be a function of soils and vegetation. These coverages did not provide significant improvements over the use of HOST, probably because the land cover types tend to correlate with soil type. As the LANDCOVER may not be stable through time (Chapter 5) it was not used further.

The philosophy behind model development was to develop relationships that were robust, not over-fitted and that were physically interpretable. To test for robustness the data set was randomly split into three sets of 80:20 percent samples. For each set, the models for a particular parameter were calibrated on the full sample and the 80% sample. The resultant models were compared to ensure that the differences in the coefficients for the independent variables were not significantly different between models. The model was re-evaluated if one or more coefficients were significantly different for more than one set. The robustness, or stability, of each model was checked by plotting the values predicted using the 80% models against the values predicted using the "100%" model for one of the 20% independent samples.

The models were developed using backwards stepwise regression with manual intervention (Systat, 1998). Continuous independent variables (transformed where appropriate) that correlated with the dependent variable were entered into the regression together with 20 sets of randomly generated numbers. The first step in the analysis was to identify significant independent variables using automatic elimination. The second step was to remove independent variables that were less significant than any of the random number variables. The random numbers which, by chance, proved to be significant were also removed at this stage.

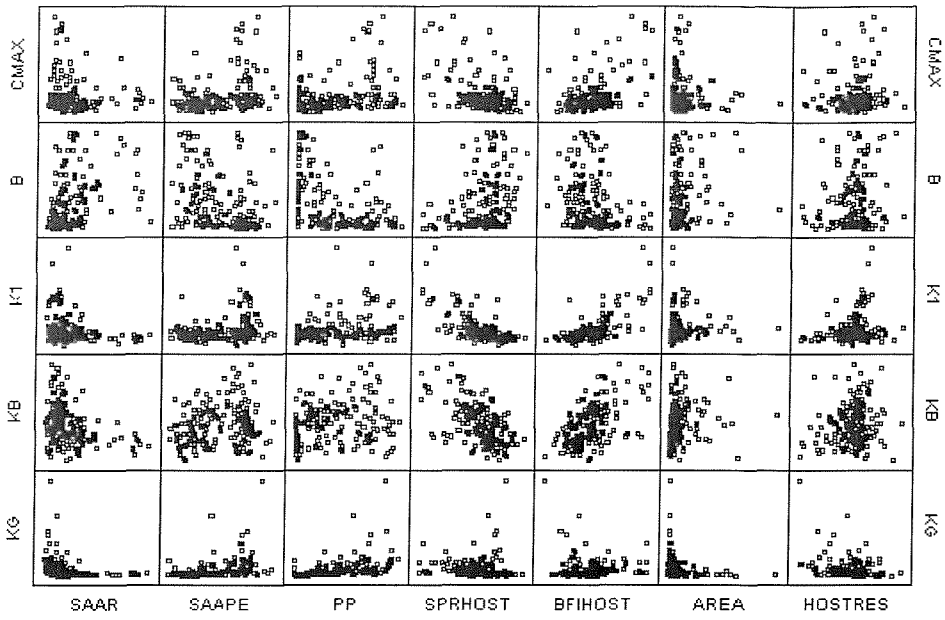



**Figure 7.5 Scatter plot matrix for the relationships between catchment characteristics and model parameters for MODA.**

One issue was that of co-linearity amongst the independent variables. The PCA of catchment characteristics and the Pearson correlation analysis identified those characteristics/descriptors that were not independent, and that could thus cause potential problems with model stability. The regression tolerance figures (one minus the squared multiple correlation between an independent variable and the other independent variables included in the model) were used to identify variables to be removed from the model during the automatic procedure (Tolerance <0.15). The tolerance values for parameters were used, in conjunction with analysis of conditions indices to further reduce co-linearity issues, and hence parameter redundancy.

Condition indices are derived by factoring the unit scaled (diagonal values of 1) $X^TX$ matrix for independent variables to yield, n, principal components, where n is the number of independent variables. The condition indices are the square roots of the ratios of the largest Eigen values to each successive Eigen values. Condition indices of greater than 15 indicate potential problems with co-linearity (Systat, 1998). Potential problems are identified from the loadings for each variable within a PC with a high condition index. The co-linearity problem is confirmed if the loadings are high for more than one variable. Instances where co-linearity was a significant problem were resolved by removing one, or more variables from the analysis. The choice as to which variable(s) to remove from the

model was conditioned by the conceptual interpretation of the influence of the variable(s) on the particular rainfall runoff model parameter being modelled and the condition indices analysis.

Outliers were identified within models based upon identifying the studentised residuals which were greater than ±2.5. Outliers are those observations that should not be included in the model because they most probably belong to different populations for which the model is not intended. Outliers were only removed if their removal was supported by appropriate physical evidence.

Where either BFIHOST or SPRHOST dominated models, alternative models were constructed using HOST directly. The final decision as to which model was finally selected for use was based on model $R^2$, SE estimates and visual inspection of plots of predicted versus observed values for the dependent variable.

The major problem encountered with fitting the HOST based models was the issue of the poor representation of certain HOST classes, both within the modelled catchments and nationally. As discussed in Chapter 5, Boorman *et al* (1995) resolved this issue for the development of the BFIHOST and SPRHOST by bounding the parameter estimates for poorly represented HOST classes using knowledge of the upper and lower bounds on BFI (0,1) and hydrologist knowledge. As the upper and lower bounds for the rainfall runoff model parameters being considered are not clearly defined, the alternative approach of Gustard *et al* (1992) was adopted in this study. In this approach poorly represented classes (with parameter estimates that are not significantly difference from zero) are amalgamated with better-represented HOST classes. This is subject to the provisos that the classes can be considered to be similar in terms of physical properties, and that the addition of the poorly represented class(es) does not unduly modify the parameter estimate for the well represented class. This is equivalent to constraining parameters for poorly represented classes to be equal to the parameter values identified for "similar" classes that are well represented.

### 7.4.2 Results

A summary of the results of the regression analysis is presented within this section. The results are presented for each model by parameter. In each case, the structure of the regression model for predicting the parameter is presented and the physical interpretation of the model structure is discussed. The full detail of the development of the regression models is presented within Appendix E. For each regression model the appendix presents a discussion of model outliers and residuals, the explained variance and an assessment of the stability of the model.

MOD A

***The maximum storage capacity, Cmax***

To obtain significant relationships between Cmax and the candidate catchment characteristics, it was necessary to express Cmax as the mean storage capacity, $\overline{C}$. This is derived by expressing Cmax as a fraction of (b+1), as discussed in Chapter 4.

The best regression relationship between $\overline{C}$ and catchment characteristics was based on regressing log ($\overline{C}$ +1) against a grouping of the fractional extents of HOST classes, called Cbar_HG. The structure of this model is given by

$$\log(\overline{C}+1) = \sum a_i \text{Cbar\_HG}_i \quad , \qquad \textbf{(7.2)}$$

where $a_i$ are the coefficents of the model. The consituent hydrogeological units for Cbar_HG classes, together with the model coefficients, $a_i$, and the standard error from the model are summarised in Table 7.5.

The grouping of HOST classes is based on the soil integrated air capacity (as a surrogate for permeability) and geological substrate. Permeable soils, particularly when underlain by permeable geologies have higher coefficients and hence higher values of $\overline{C}$. This is conceptually correct as these soils will have high infiltration rates and will have a greater storage capacity. Conversely, thin impermeable soils, or thin soils underlain by impermeable geologies will have low storage capacities.

**Table 7.5 The Cbar_HG classification and model coefficients**

| Cbar_HG | Hydrogeological units | Coefficient (a) |
|---|---|---|
| 1 | Chalk, chalk drift | 0.026 |
| 2 | Oolitic limestone, soft magnesian | 0.019 |
| 3 | Blown sand, gravels | 0.023 |
| 4 | Colluvium, coverloam, gravel, loamy drift | 0.019 |
| 5 | Very soft massive clays | 0.02 |
| 6 | Hard rock | 0.019 |
| 7 | Soft bedded clays, loams, till, shales | 0.02 |
| 8 | Soft shales, siltstones, bedded clays/loams, clay and flints | 0.017 |
| 9 | Weathered intr/meta rock, raw peats, eroded peats | 0.008 |
| 10 | Soft sandstone | 0.027 |
| 11 | Weathered intr/meta rock, hard limestone and sandstone | 0.023 |
| 12 | Weathered intr/meta rock, colluvium, coverloam, loamy, drift gravels/loams, sandstone | 0.016 |
| 13 | Weathered intr/meta rock, chalky drift, loams Earthy peats, Shattered rock, alluviums, covei loam, chalky drift | 0.012 |
| | adjusted $R^2$=0.988 | S.E.= 0.199 |

The model is essentially a weighted average of percentage coverages of the HOST groupings, with the weight given by the coefficient estimate. This is conceptually attractive, as $\overline{C}$ is an average value of storage capacities. The logarithm of ($\overline{C}$ +1) is used within the regression so that the model can predict a zero value of $\overline{C}$ for a catchment with zero fractional extents. This cannot occur in practice as it implies a zero catchment area, but it is mathematically correct. The adjusted $R^2$ value is very high, a function of the fact the model has a zero intercept and is therefore not constrained to pass through the mean of the data. Of more importance is the Standard Error (SE) of 0.199. Once antilogs are taken, this equates to a Factorial Standard Error of 1.58, that is (given the number of data points is high) the predicitve accuracy of the model is approximately ±58% at the 68% confidence level.

***The Pareto shape parameter, b***

The optimal model for the Pareto shape parameter b was a linear model that related b to the fractional extent of a six-class grouping of HOST classes, b_HG, given by

$$b = \sum b_1 b\_HG_1 \ , \qquad \textbf{(7.3)}$$

where $b_i$ are the model coefficients. The grouping, regression coefficients and model fit statistics are presented in Table 7.6 for this model.

**Table 7.6 The b_HG classification and model coefficients**

| b_HG | Hydrogeological units | Coefficient (b) |
|---|---|---|
| 1 | Weathered intr/meta Rock, sandstone, colluvium, gravels/loams | 0.023 |
| 2 | Weathered intr/meta Rock, hard limestone and sandstone Blown sand, gravels Colluvium, coverloam, gravel, loamy drift Soft shales, siltstones, bedded clays/loams, clay and flints Weathered int/meta Rock, colluvium, coverloam, loamy drift | 0.007 |
| 3 | Chalk, chalk drift Oolitic limestone, soft magnesian Weathered intr/meta rock, chalky drift, loams Shattered rock, alluviums, cover loam, chalky drift Very soft massive clays Soft bedded clays, loams, till, shales Soft sandstone | 0.004 |
| 4 | Weathered int/meta Rock, raw peats, eroded peats Earthy peats | 0.052 |
| 5 | Hard rock | 0.014 |
| | Adjusted $R^2$=0.658 SE= 1.046 | |

The regression model fit was very poor, this is probably due in part to the covariance between Cmax and b. The standard error is 1.046, giving an approximate predictive accuracy of ±1 at the 68% level. The mean and standard deviation of the raw observed data set are 1.2 and 1.3 respectively. This indicates that the model does not give a major improvement over just using the mean, although it should be noted that the parameters are all significant.

The HOST grouping was obtained using the grouping strategy that was used for the mean storage capacity, $\overline{C}$. Classes corresponding to thin soils overlying hard rock substrate have a high coefficient value of b, which implies that the storage distribution will be skewed towards the shallow soil storage capacities. The model predicts that the distribution of soil capacities on permeable soils, particularly when overlying permeable substrates, will be skewed towards the maximum soil capacity (low value of b). The implications of this are that the impermeable soils will start to generate direct runoff earlier than permeable soils for a given precipitation event. Furthermore, significant soil moisture deficit will occur more frequently in these lower storage soils. This is conceptually correct, and it is consistent with the observations of model behaviour for MODA, discussed within Chapter 6.

***The loss model drainage constant, Kg***

The optimal model identified for the drainage constant was one relating the logarithm of (Kg+1) to the logarithms of PP, LDP and HOSTRES. The form of the model is given by

$$\log(Kg+1) = 8.593\log(PP+1) - 0.306\log(LDP+1) - 4.257\log(HOSTRES+1) + 3.341 \quad \textbf{(7.4)}$$

$R^2 = 0.58$ S.E. = 0.23

The model explains 58% of the variance and the S.E. for the model equates to a F.S.E. of 1.7. The predictive accuracy of the model is therefore approximately ±70% at a confidence level of 68%.

The model implies that Kg will be larger and therefore limiting drainage for catchments that have a greater tendency to build up soil moisture deficits, as represented by PP. The negative coefficient for LDP indicates Kg is smaller in large catchments, which implies that the drainage, and hence base flow, is potentially greater in larger catchments. This is an indication that the model is representing the averaging effects of larger catchments by increasing the drainage from the soil store, and thus the fraction of effective rainfall routed through the slow flow reservoir. Kg decreases with increasing HOSTRES. A high value of HOSTRES indicates that the catchment has a higher BFI (and hence base flow) than would be expected based upon anticipated SPR, again this is intuitively correct.

***The slow flow routing reservoir time constant, Kb***

The optimal, stable model identified for this parameter was one that relates Kb to BFIHOST and the logarithm of AREA. The structure of the model is given by:

$$Kb = 1389\,BFIHOST + 170\log(AREA+1) - 260 \quad \textbf{(7.5)}$$

$R^2 = 0.51$ S.E. = 298

The model has a relatively low adjusted $R^2$ and significantly under predicts large values of Kb. The model predicts higher values of Kb for both high base flow (BFI) and larger

catchments. This is conceptually correct, permeable systems are high storage systems in which stream flow is dominated by release of water from groundwater, with a generally small direct runoff component. As a result, rates of recession rates are low in these catchments, which is reflected in the rainfall runoff model by higher values of Kb. In larger catchments, both hill slope and groundwater routing will become more damped and channel routing and storage effects may start to become important. One effect of this would be to reduce both the variance of stream flow data and the rates of changes within stream flow. Therefore, a relationship with area for the routing time constants is conceptually acceptable.

***Quick flow routing reservoir time constant, K1***

The best, stable model fits were obtained for a model that related Log (K1+1) to the fractional extents of a grouping of HOST classes (K1_HG). The fit obtained with the model was similar to the fit of a model based on BFIHOST and LOG (AREA+1), but the HOST-based model gave a better fit at low values of K1. The structure of the model is given by

$$\log(K1+1) = \sum c_i K1\_HG_i \quad , \qquad \textbf{(7.6)}$$

where $c_i$ are the coefficents of the model. The consituent hydrogeological units for K1_HG classes, together with the model coefficients, $c_i$, and the fit statistics for the model are summarised in Table 7.7. The grouping of HOST classes is based on permeability and the physical charatceristics of the substrate geology. Permeable soils that are underlain by permeable geologies have higher coefficients. Overland flow that reaches the stream, and or minor channels will not normally be generated within very permeable systems. The direct runoff that there is will be dominated by inter-flow giving a more damped response, and hence higher values of K1. This is conceptually correct. The FSE for the model is 1.57, that is the predicitive accuracy of the model is approximately ±57% at the 68% level. Analysis of residuals demonstrated that the model has a tendency to over predict very low values of K1.

**Table 7.7 The K1_HG classification and model coefficients**

| K1_HG | Hydrogeological units | Coefficient (c) |
|---|---|---|
| 1 | Chalk, chalk drift | 0.023 |
| 2 | Oolitic limestone, soft magnesian | 0.022 |
| 3 | Blown sand, gravels | 0.021 |
| 4 | Very soft massive clays | 0.016 |
| 5 | Hard rock | 0.019 |
| 6 | Soft bedded clays, loams, till, shales | 0.018 |
| 7 | Soft shales, siltstones, bedded clays/loams, clay and flints | 0.017 |
| 8 | Weathered intr/meta. Rock, raw peats, eroded peats | 0.015 |
| 9 | Soft sandstone | 0.018 |
| 10 | Weathered intr/meta. Rock, hard limestone and sandstone | 0.019 |
| 11 | Weathered intr/meta. Rock, sandstone, colluvium, loamy drift, gravels/loams | 0.015 |
| 12 | Earthy peats<br>Colluvium, coverloam, gravel, loamy drift<br>Weathered intr/meta. Rock, chalky drift, loams | 0.021 |
| | adjusted $R^2$=0.99 | SE=0.195 |

MODB

The MODB parameters to be regionalised were Cmax, K1 and Kb. The regression models for the MODB parameters are presented in the subsequent sub-sections.

***The maximum storage capacity, Cmax***

The modelling of Cmax proved to very problematical. The optimal model derived related the logarithm of (Cmax+1) to the logarithms of (PP+1) and (DPLBAR+1), and is given by

$$\log(Cmax+1) = 5.58\log(PP+1) - 2.61\log(DPLBAR+1) + 2.17 \quad . \qquad \textbf{(7.7)}$$

$$R^2 = 0.36 \qquad S.E. = 0.296$$

The fit of the model is poor, explaining 36% of the variance and has a FSE of 1.98. The physical interpretation of the model is not clear, but may be related to the role of Cmax in controlling the evaporation rate. A uniform distribution of soil storage capacities is assumed for MODB. The relationship between (precipitation-evaporative losses) and outflow from the soil store is therefore quadratic (Chapter 4). Cmax controls the gradient of this relationship and fixes the storage capacity at which the catchment is fully saturated.

Cmax also directly controls the relationship between the evaporation rate and simulated Soil Moisture Deficit (SMD). However, Cmax can only influence evaporation rates if the summer evaporation demand exceed the rainfall inputs (hence generating an SMD). Cmax is probably only really identifiable under these circumstances. PP is a measure of the potential for significant SMDs to build up in a catchment. The regression model predicts that Cmax increases as PP increases. If PP is high the model will certainly generate a soil moisture deficit, and will potentially result in a reduced evaporation rate. Given the formulation of the evaporation function, this tendency will be offset by an increased value of Cmax. The relationship between Cmax and PP may well be an artefact of the rainfall runoff model structure. The dependency on DPLBAR implies that Cmax is larger for catchments with larger mean drainage path lengths. This characteristic is high for both large catchments and catchments that tend to be linear. This is difficult to explain, but may be associated with the longer runoff concentration times within these catchments.

***The routing reservoir time constants, K1 and Kb***

Under ideal circumstances the time constants obtained for MODA would be identical to those obtained for MODB, as they were un-correlated with the loss module parameters. However, the values are influenced by the differences in the partitioning mechanism for effective rainfall between the two models. In practice, they were identified as being very comparable. The optimal, stable regression model structures identified are very similar to those obtained for MODA and are given by

$$Kb = 1872\,BFIHOST + 206\log(AREA + 1) - 491 \qquad \textbf{(7.8)}$$

$R^2 = 0.50 \quad S.E. = 306$

and

$$\log(K1 + 1) = 1.279\,BFIHOST + 0.224\log(LDP + 1) + 0.812 \qquad (7.9)$$

$R^2 = 0.52 \quad S.E. = 0.182$

The model for Kb is identical in structure to that obtained for Kb with MODA. The coefficients are different and the model is a slightly poorer fit. This may a consequence of

the cruder treatment of the partitioning of effective rainfall in MODB. The model for K1 uses a log-transformed representation of K1, which is a function of BFIHOST and Log (LDP+1). This formulation gave similar results for K1 with MODA, and was only a slightly poorer fit than the HOST based model finally selected for MODA. The same formulation applied to MODB gives a predictive accuracy of ±52%, which is an improvement of ± 5% over the predictive capacity of the HOST-based K1 model for MODA The use of LDP gave a significant advantage over the use of AREA. Whilst these are strongly correlated, they are different characteristics. The relationship indicates a damping of direct runoff response to precipitation events in long drainage path catchments. This is conceptually attractive as it represents catchment and channel routing considerations that are not explicit within the model structure.

## 7.5 SUMMARY OF THE REGIONALISATION STUDIES

The evaluation of the calibrated fit of both model configurations within the full catchment set demonstrated that the quality of model fits was similar across the catchments. The majority of model fits were very good; MODA is better at predicting mean flow and hence conserving mass and MODB is better at predicting low flows. Both models were stable when applied to a evaluation period, this was probably aided by the evaluation period being generally wetter than the calibration period for the majority of catchments, thus making the modelling of soil moisture behaviour less critical. The model fits obtained for MODB are generally more stable that those obtained for MODA, this indicates that MODA may be over specified.

Principal Components Analysis was successfully applied to the physiographic and climatic catchment characteristics/descriptors for the catchments where MODA was successfully calibrated. The analysis demonstrated that the variation in the data set was dominated by the climatological, altitude and slope characteristics. This dominance is a reflection of the strong correlation relationships between these variables and the large variation across the UK within these characteristics.

Relationships between model parameters and catchment characteristics/descriptors were developed using a multivariate regression approach for both MODA and MODB, and a ROI approach for MODA. The ROI approach for regionalising MODA, in which the internal structure of the individual parameter vectors is retained, was found to be not as effective as the use of the regression-based approach. Good or acceptable regression relationships were identified for all model parameters with the exception of the Pareto shape parameter (b), in the case of MODA and Cmax, in the case of MODB. These are both parameters for the soil moisture store. The parameter b controls the distribution of soil storage capacities within MODA, and hence influences the generation of direct runoff, evaporation rates and drainage rates from the base of the store. Cmax is the maximum storage capacity per unit area within the store. In MODB this parameter controls the sensitivity of the rainfall runoff relationship to soil moisture status, and the relationship between evaporation rates and soil moisture deficits within the soil store.

The uncertainties in the relationships between model parameters and catchment characteristics are a function of the physical reality of the structure of the rainfall runoff model, the errors in input data, calibration errors, errors in catchment characteristic data and errors, or assumptions in structure of the relationships themselves. In the context of this study it was prerequisite that the calibrated rainfall runoff model could close an effective water balance over the calibration period, and that this closure was stable when the model was applied to an independent evaluation period of record. The assumption of a closed water balance is very sensitive to errors in climatic data (particularly precipitation), an effective estimate of contributing area and the assumption that flows do not bypass the gauging station. The analysis presented in Chapter 6 demonstrated that, during calibration, the loss module parameters compensate for these types of error through manipulation of both the evaporation rate and the time that water is retained in the soil store. This is one major reason for the poor relationships between b and catchment characteristics in MODA, and Cmax and catchment characteristics in MODB. A second issue for MODA is the strong covariance between Cmax and b. This issue was demonstrated by the fact that significant relationships between Cmax for MODA and catchment characteristics could not be identified, and yet a good relationship for predicting $\overline{C}$ (which is a function of Cmax and b) was identified.

# 8 Assessment of the performance of the regionalised rainfall runoff models

## 8.1 INTRODUCTION

The predictive capacity of the regression models for estimating model parameters is of importance, however it is the predictive performance of the rainfall runoff models, when run with the estimated parameters, that is ultimately of interest. The performance of the regionalised models was assessed by:

- evaluating the fits obtained within the case study catchments, considered in Chapter 6, with respect to the observed flows and the simulated flows obtained with the calibrated model parameters;
- evaluating general patterns in the quality of model fit obtained using the regression based parameter estimates across all of the catchments for which calibrated parameters were obtained;
- developing a classification scheme for classifying model simulations in the context of their applicability for use within water resource assessments;
- comparing the performance of the regionalised model fits within two of the Anglian catchments used in the model evaluation studies presented in Chapter 3;
- comparing the performance of the regionalised model fits with those obtained by transposing local data from analogue catchments.

The evaluation with the specific case study catchments is presented in Section 8.2 and the evaluation across all catchments is presented in Section 8.3. A classification of results for water resources assessment is presented within Section 8.4. The comparisons with the Anglian modelling studies are presented in Section 8.5. A summary of the results from these sections is presented in Section 8.6. The use of data transposed from analogue catchments, and a comparison of this approach with the use of a regionalised model is presented within Chapter 9.

## 8.2 THE PERFORMANCE OF REGRESSION MODELS WITHIN THE SPECIFIC CASE STUDY CATCHMENTS

The locations and characteristics of the case study catchments are discussed in detail within Section 6.5.1 of Chapter 6. As discussed in Chapter 6, the four case study catchments represent the following climatological and hydrological permutations:

- dry- permeable:- The Bure at Ingworth (34003);
- dry-impermeable:- The Pinn at Uxbridge (39098);
- wet- permeable:- Sydling Water at Sydling St. Nicholas (44006);
- wet-impermeable:- The Gelyn at Cynefail (67010).

**Table 8.1 Model parameters and fit statistics for the case study catchments**

| Gauge | Model parameters | | | | | | Quality of model fit | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | gamma | Cmax | b | Kg/Beta | K1 | Kb | BIAS | EFF | BEQ95 | BIAS | EFF | BEQ95 |
| *MODA* | | *Calibrated Parameters* | | | | | *Calibration Period* | | | *Evaluation Period* | | |
| 34003 | 2.03 | 793 | 1.07 | 6374 | 57 | 1639 | 2 | 0.53 | -20 | -1 | 0.54 | 6 |
| 39098 | 1.08 | 373 | 0.95 | 51601 | 26 | 144 | 1 | 0.68 | -11 | -7 | 0.43 | -6 |
| 44006 | 1.67 | 433 | 0.03 | 7365 | 384 | 502 | 0 | 0.60 | 24 | 3 | 0.72 | 28 |
| 67010 | 1.78 | 110 | 3.04 | 762 | 30 | 496 | -3 | 0.73 | -6 | -2 | 0.76 | 4 |
| *MODA* | | *Estimated Parameters* | | | | | | | | | | |
| 34003 | 2.03 | 241 | 0.66 | 4108 | 111 | 1201 | 4 | 0.25 | -47 | 3 | 0.34 | -29 |
| 39098 | 1.08 | 139 | 0.40 | 13623 | 39 | 247 | 23 | 0.52 | 28 | 15 | 0.62 | 51 |
| 44006 | 1.67 | 447 | 0.4 | 2410 | 165 | 1153 | 24 | 0.73 | 64 | 20 | 0.72 | 57 |
| 67010 | 1.78 | 66 | 3.33 | 981 | 40 | 284 | -2 | 0.73 | -20 | -1 | 0.77 | -30 |
| *MODB* | | *Calibrated Parameters* | | | | | *Calibration Period* | | | *Evaluation Period* | | |
| 34003 | 0 | 251 | 1 | 0.78 | 112 | 1770 | 0 | 0.17 | -47 | 0 | 0.29 | -22 |
| 39098 | 0 | 386 | 1 | 0.17 | 32 | 398 | 0 | 0.73 | -22 | -4 | 0.68 | -5 |
| 44006 | | | | | | | | | | | | |
| 67010 | 0 | 37 | 1 | 0.25 | 22 | 492 | 2 | 0.78 | -9 | 3 | 0.81 | 9 |
| *MODB* | | *Estimated Parameters* | | | | | | | | | | |
| 34003 | 0 | 242 | 1 | 0.78 | 132 | 1427 | 1 | 0.00 | -55 | 1 | 0.14 | -33 |
| 39098 | 0 | 267 | 1 | 0.17 | 20 | 157 | 18 | 0.58 | -55 | 11 | 0.63 | -48 |
| 44006 | | | | | | | | | | | | |
| 67010 | 0 | 88 | 1 | 0.25 | 21 | 216 | -1 | 0.78 | -58 | 0 | 0.81 | -36 |

The models, MODA and MODB, were run over the calibration and evaluation periods for the case study catchments using the regression based parameter estimates. The results obtained using these regression based parameter estimates were compared to those obtained with the calibrated model parameters over both the calibration and evaluation period.

The model parameters and goodness of fit measures obtained are presented for both models within Table 8.1. For each model, and each catchment this table presents: the selected calibration parameters used in the regression modelling; the parameters estimated using the regression models and an assessment of the goodness of model fit over the calibration and evaluation periods - as represented by the BIAS, EFF and BEQ95 statistics. The results obtained for the individual catchments are discussed in the following sub-sections.

### 8.2.1 Catchment 34003

Within this catchment, the fit of MODA using calibrated parameters is better than that obtained for MODB using the calibrated parameters over both the calibration period and the evaluation period. The EFF values for MODB are very low indicating a poor time series fit for this model. The fit of MODA using the calibrated parameters, as judged by the goodness of fit measures, is better than that obtained using the regression-based parameters over both the calibration and evaluation periods. This pattern is repeated for MODB. The fit obtained using the calibrated parameters for MODB is approximately equivalent to that obtained for MODA using the regression-based parameters. Both models adequately close a water balance with both calibrated and estimated model parameters over both the calibration and evaluation periods. Simulated and observed hydrographs for the period 1987-88 are presented in Figure 8.1. This period falls within the calibration period used for the catchment. The simulated hydrographs are those obtained for both models using the calibrated and regression estimated parameter estimates. The corresponding soil moisture status of the loss module is presented within Figure 8.2 for the same period. In this figure, the depth of water held in storage within the catchment per unit area is expressed as a fraction of the maximum storage depth, Smax, for each model configuration. If this ratio is less than 0.5, then the evaporation rate will be reduced to below the potential rate, as discussed within Section 4.2.2 of Chapter 4. As is discussed within Chapter 4, Smax is equal to Cmax/(b+1) for MODA, and is equal to Cmax/2 for MODB.

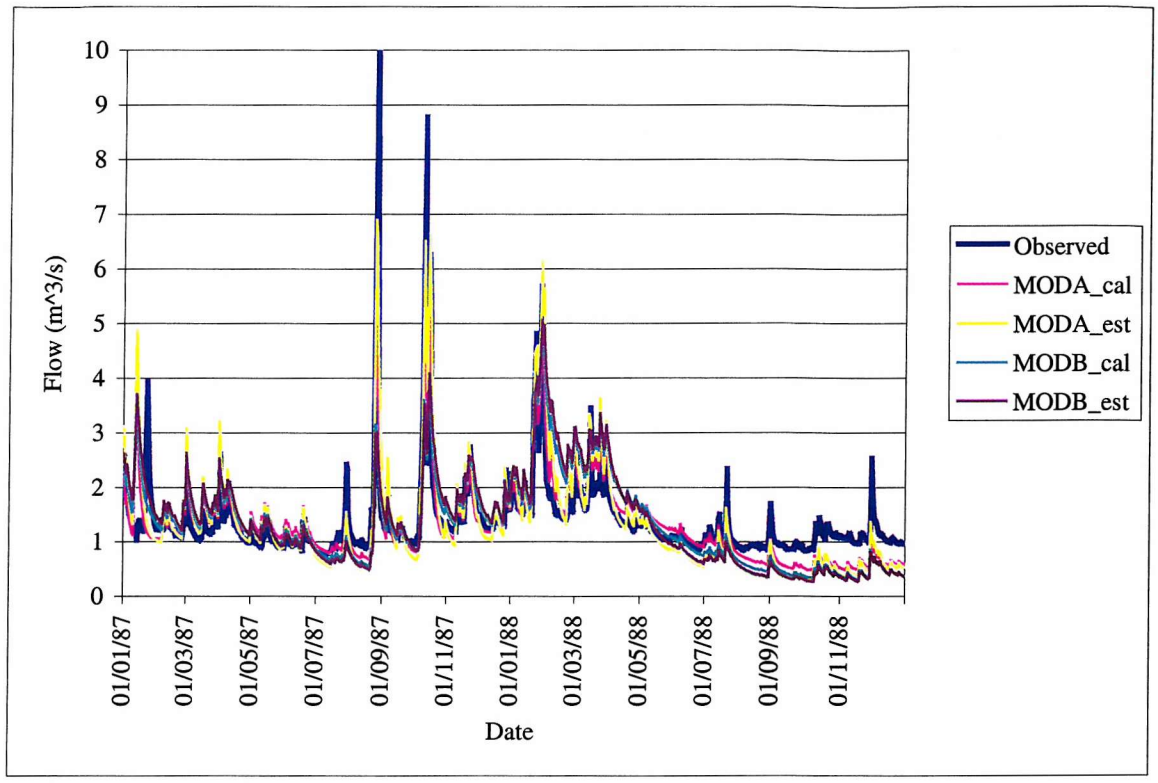



**Figure 8.1 Example observed and simulated hydrographs for 34003.**

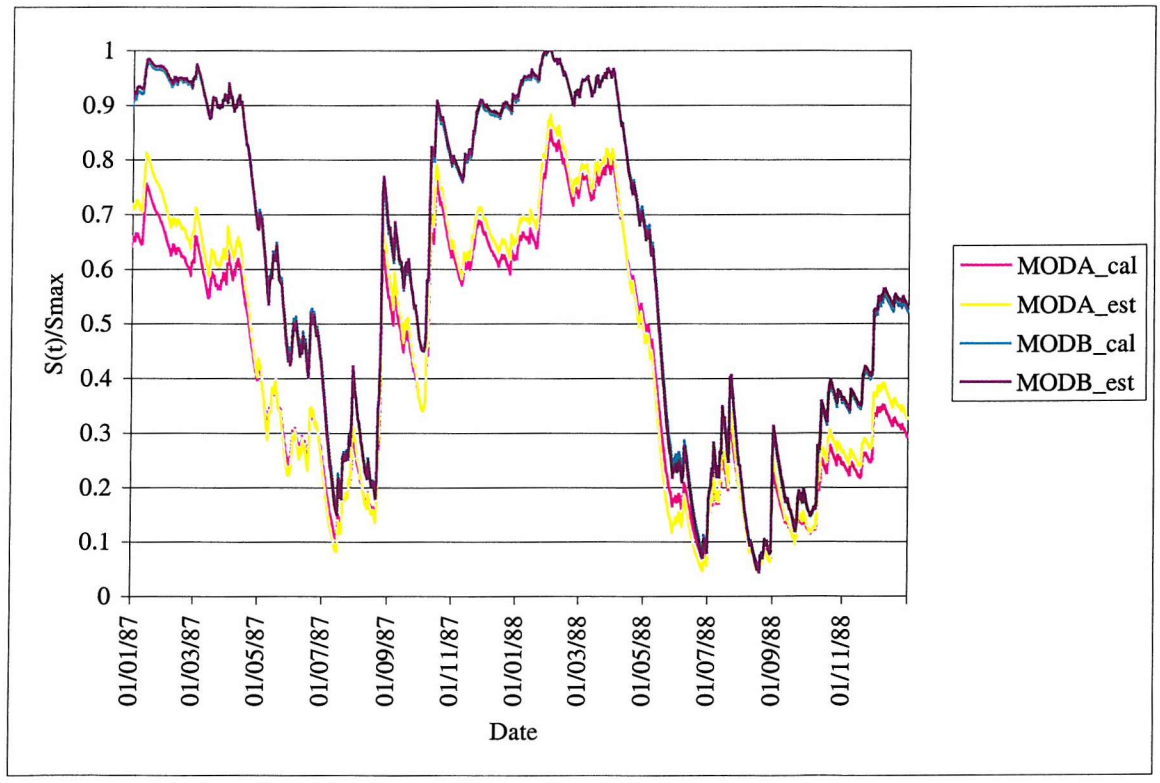



**Figure 8.2 Example graphs of simulated soil moisture status for 34003.**

MODA adequately simulates the observed hydrograph for flows above $1m^3s^{-1}$ for both calibrated and estimated parameter vectors. MODB tends to over estimate flows in the wetter periods, probably due to an incorrect partitioning of effective rainfall on the basis of BFIHOST. This catchment is permeable with a high infiltration capacity, and thus the direct runoff component will be low for all but the largest rainfall events. The over-estimation of higher flows (but not extreme events) by MODB probably accounts for the poor EFF values observed for the model for both calibrated and regionalised model configurations.

The soil moisture behaviour is similar for the two MODA model configurations. The two MODB configurations also have a similar soil moisture behaviour, however MODA predicts that the loss module is never more than 90% full whereas MODB predicts that the catchment can become fully saturated within the winter months. This behaviour is a consequence of the much lower values of Smax for the MODB configurations. Even though the catchment is a low rainfall catchment, it would be anticipated that the catchment would be saturated at points within the winter months. On this basis, the MODB simulation of soil moisture behaviour is more realistic. The model predicts that SMDs are at a minimum at the end of the winter recharge season (normally March), and that evaporation limiting SMDs occur within the summer months. As can be seen from the stream flow hydrographs, the observed and simulated runoff response to rainfall events is very low during these periods of high SMD, as would be expected. This behaviour is consistent with the permeable nature of the catchment.

There appears to be a sustained observed base flow within the catchment of $1m^3s^{-1}$. This may be a natural base flow; or it may be a consequence of a discharge that was not accounted for, or it may be a hydrometric error. Both models are unable to simulate this base flow but are both able to simulate the catchment response to rainfall events when the observed flows are above $1m^3s^{-1}$. The inability to model the observed flow regime at low flows accounts for the large negative systematic errors at Q95 flows observed for all configurations over the calibration period. The evaluation period was considerably wetter than the calibration period, with generally higher base flows, and thus the systematic error at Q95 is much less for both models, and both sets of parameters. MODA, when used with the calibrated parameter vector, consistently over estimates Q95 over the evaluation period.

Within the example hydrograph, it can be seen that the errors in base flow are generally larger for MODB using estimated parameters and smaller for MODA using estimated parameters. This is confirmed by the BEQ95 statistic for the calibration period.

### 8.2.2 Catchment 39098

Within this catchment, the fit of MODA using the calibrated parameters is better than that obtained using the regression based parameters over both the calibration and evaluation periods. This pattern is repeated for MODB. The calibrated fit of MODA is marginally poorer than that obtained for the calibrated form of MODB over both the calibration and the evaluation periods. This pattern is repeated for the estimated parameters, with the MODB model performing better than MODA over both periods. When using the regionalised model parameters, both models over estimate mean flow (positive BIAS) by more than 10% over both the calibration and evaluation periods. This tendency to over predict is greater for MODA than MODB, particularly over the evaluation period. The tendency to over predict is linked to the significant under estimation of Cmax by the regression models for both models. Significant SMD will build up in this dry catchment, and a small value of Cmax results in evaporation rates being reduced to below the potential rate more frequently and to a greater extent leading to an over-prediction of mean flow.

Simulated and observed hydrographs for 1987-88 are presented in Figure 8.3 for both models using the calibrated and regression based parameter estimates. This period falls within the calibration period used for the catchment. The corresponding soil moisture status of the loss module is presented within Figure 8.4 for the same period. Both models simulate the catchment response to rainfall events well. During recession periods the simulated flow for MODA is slightly too high when using the regionalised parameters. This is a consequence of the underestimation of Kg, which results in more of the effective rainfall passing through the slow flow routing reservoir. The predicted recession rate at low flows is too high for MODB when using the regionalised parameters. This is a consequence of the under estimation of the time constant for the slow flow routing reservoir and results in an under–estimation of Q95 flows. These flows are small for this catchment, and thus a large percentage error does not equate to a large volumetric error. The performance of the two models, when used with the calibration parameters is very similar, and is close to the observed behaviour during recession periods.

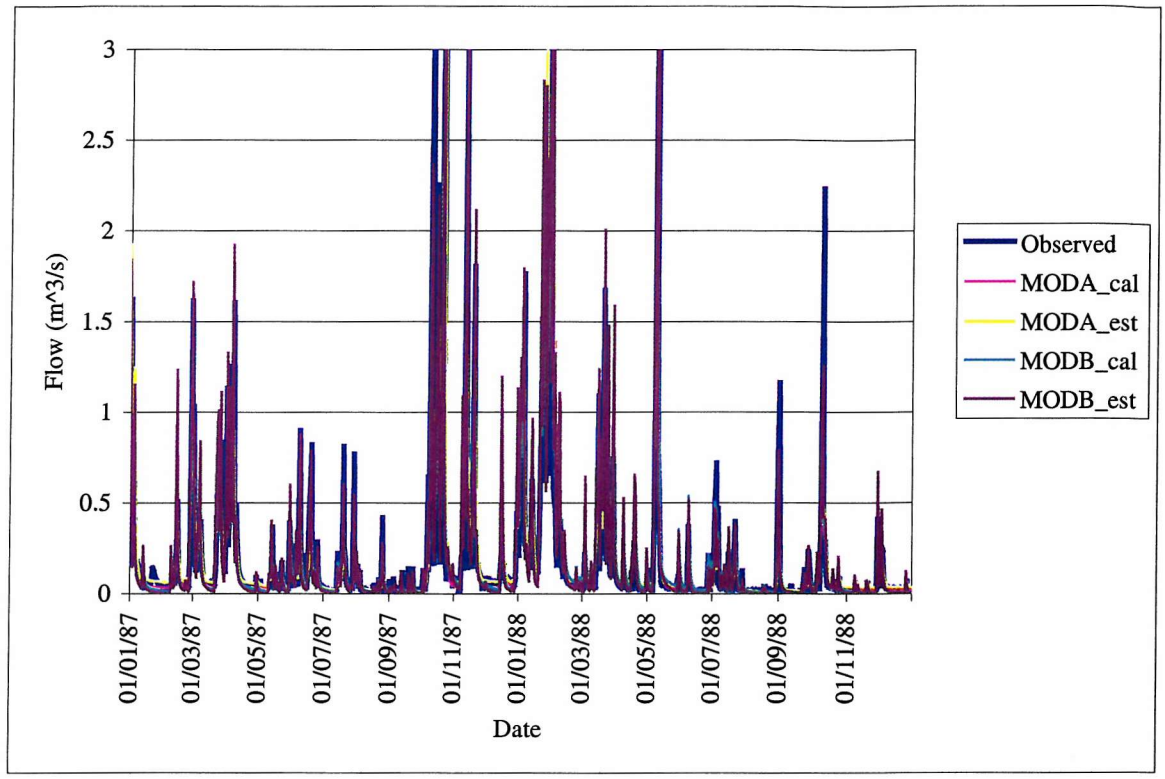



**Figure 8.3 Example observed and simulated hydrographs for 39098.**

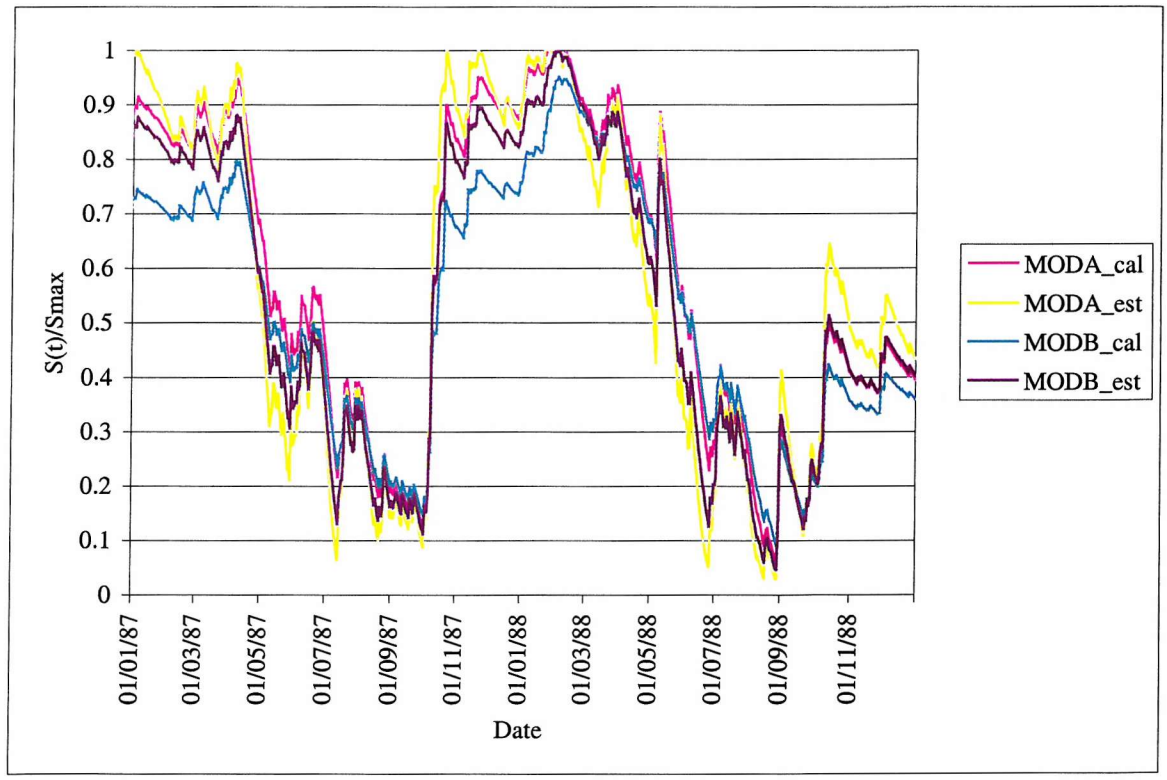



**Figure 8.4 Example graphs of simulated soil moisture status for 39098.**

The soil moisture behaviour is similar for all model configurations. The variation is slightly lower for MODB when run with the calibrated model parameters, this is a consequence of the greater storage capacity for the loss module within the configuration, compared to the others. Soil moisture deficits exist for the majority of the time; the catchment is rarely fully saturated, and is only saturated within the winter months when evaporation demand is lowest. Evaporation-limiting SMDs occur within the summer months. As can be seen from the stream flow hydrographs the observed and simulated runoff response to rainfall events is lower during these periods of high SMD, as would be expected. This behaviour is commensurate with the impermeable nature of the catchment and the low rainfall regime of the catchment. There is evidence that evaporation limiting SMDs are greater for the regionalised model forms which results in an over estimate of mean flow, as discussed.

### 8.2.3 Catchment 44006

There are no results for MODB for this catchment as the model was unable to close an adequate water balance during the calibration period. Simulated and observed hydrographs for 1987–88 are presented in Figure 8.5. These simulated hydrographs are those obtained for MODA when using the calibrated and regression estimated parameter estimates. This period again lies within the calibration period used for the catchment. These hydrographs demonstrate that the model cannot simulate the catchments quick response to rainfall events when it is run with the calibrated parameters, even though the goodness of fit statistics are good over the calibration period. The simulation of the base flow is quite reasonable.

The corresponding soil moisture status of the loss module is presented within Figure 8.6 for the same period. The calibrated parameter estimate for Kg is relatively high, indicating that a significant proportion of runoff is being routed through the quick flow reservoir. The time constant for K1 is high and the time constant for Kb is relatively low for a permeable catchment. The simulated base flow is therefore controlled by the outflow from both reservoirs. The relative proportion of the effective rainfall passing through the two stores is controlled by the BFI calculated from the flow record during calibration (Chapter 5). This indicates that the calculated BFI is not a true measure of the partitioning of effective rainfall within this catchment, leading to a calibration error. This may well be a

consequence of the fact that the catchment at 12km$^2$ is very small and that the calculation of BFI is influenced by local channel controls on stream flow generation.

The estimated parameters yield a more realistic simulation of the quick flow response of the hydrograph, however the base flow component of the simulation is too high. The estimated parameter estimates for Kg, K1 and Kb are similar to the calibrated values obtained for similar permeable catchments. The drainage is high (small Kg). The slow flow time constant is high giving a very damped base flow response, and K1 is small thus providing the quick response to rainfall.

The large calibrated values of Kg and Cmax, and the low value of b will result in water being retained within the loss module and hence lost through evaporation, as discussed in Chapter 6. This enables a water balance to be closed within this catchment. The lower regression based estimate of Kg and higher estimate of b will result in the loss module having a lower residence time, and thus less water will be evaporated off at the potential rate – resulting in an over-estimate of base flow and mean flow. This strongly indicates that the calibration procedure was compensating for errors in the input data that result in a violation of the water balance assumption. The simulated soil moisture behaviour observed for this catchment confirms this. The calibrated form is retaining more water within the loss module than the regionalised form. The regionalised form predicts that evaporation-limiting SMDs occur for the majority of the time. Given the relatively high rainfall regime for the catchment, this is incorrect. This catchment is a small, groundwater fed catchment and the most likely explanation for this behaviour is that the topographic divide is an underestimate of the contributing groundwater catchment, leading to an underestimate of the true precipitation inputs to the catchment. This also explains why MODB could not close a water balance within this catchment. MODB has a fixed partitioning of effective rainfall and uniform distribution of soil moisture stores and thus does not have the same flexibility for closing an adequate water balance on erroneous input data.

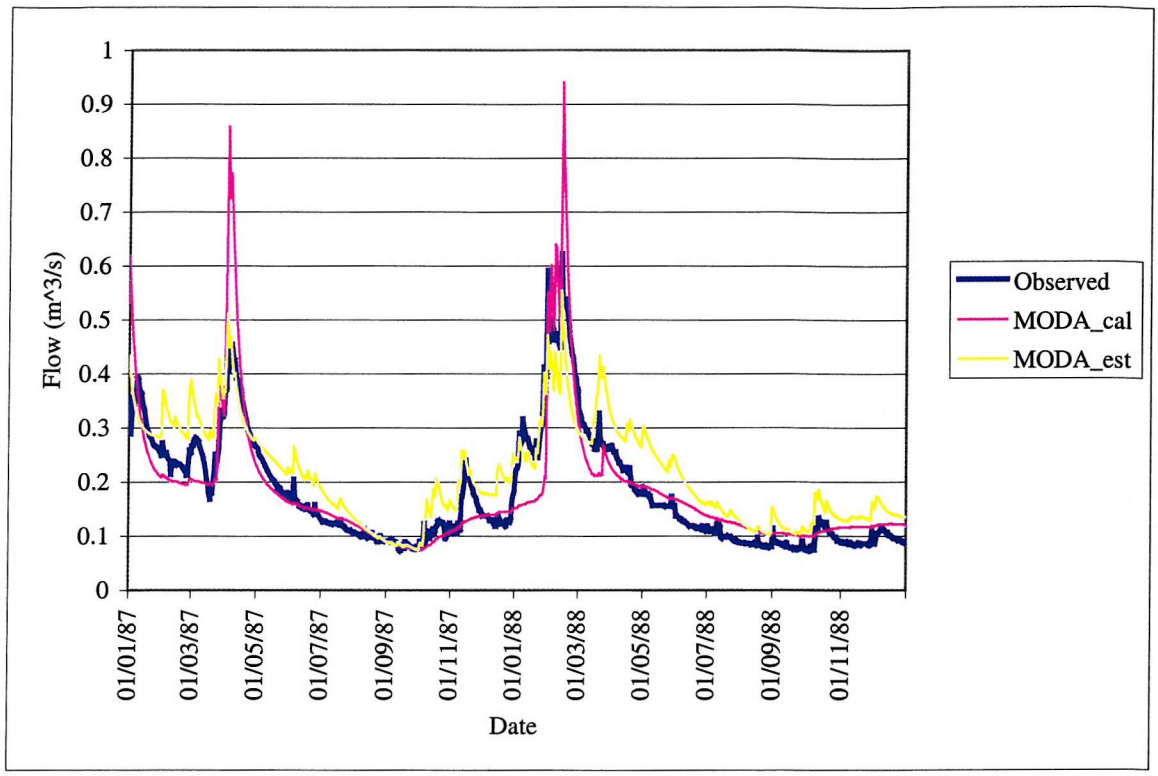



**Figure 8.5 Example observed and simulated hydrographs for 44006.**

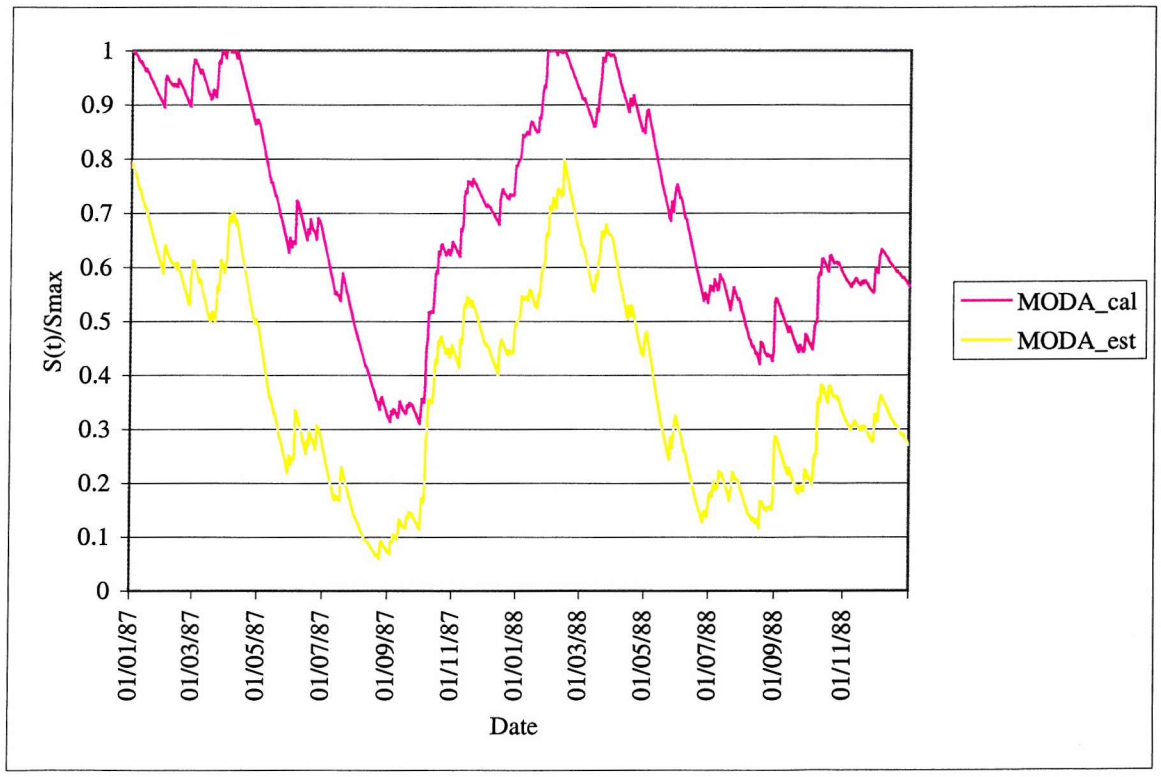



**Figure 8.6 Example graphs of simulated soil moisture status for 44006.**

### 8.2.4 Catchment 67010

All model configurations are effective in closing a water balance within this catchment over both the calibration and evaluation periods. The time series fit, as assessed using EFF, is also very good for all simulations. Both models are unbiased at the Q95 flow when the calibrated model parameter vectors are used and tend to under estimate Q95 when used with the regression based parameter estimates. The Q95 flows within this catchment are very low and, from a volumetric perspective, the errors at Q95 are still very small when the regression-based parameters are used. This tendency for the regression based parameters to result in an under prediction of low flows can be seen within the example hydrographs for this catchment, Figure 8.7. When the models are run with the regression-based parameters, the base flow recession rates are too high. This is a consequence of the under estimation of the slow flow reservoir time constants, Kb. For MODA the calibrated and estimated values of Cmax and Kg are small and the calibrated and estimated values of b are large. The drainage rate for the loss module is low, despite the small value of Kg. The reason for this is that the maximum soil moisture store depth is low, and the distribution of soil moisture stores is low. This results in a greater proportion of the effective rainfall forming a quick stream flow response to rainfall events in this catchment.

The corresponding soil moisture status of the loss module is presented within Figure 8.8 for the example period. The soil moisture behaviour is similar for all model configurations. All model configurations predict that soil moisture deficits can build up very rapidly within the catchment. This is a function of the low values of Smax predicted for the catchment, which in turn is a function of the highly impermeable nature of the catchment. Evaporation limiting soil moisture deficits are only sustained during summer months, and rarely last for more than a few weeks. Examples are July 1979 and the May 1980. As can be seen from the stream flow hydrographs, the observed and simulated runoff response to any rainfall is negligible during these periods of high SMD, as would be expected. Inspection of the rainfall records shows that these periods were both low rainfall periods (particularly May 1980) and any rainfall that did fall was retained within the soil. The catchment is commonly fully saturated within the winter months when evaporation demand is lowest. The model behaviour is commensurate with the highly impermeable nature of the catchment and the high rainfall regime of the catchment.

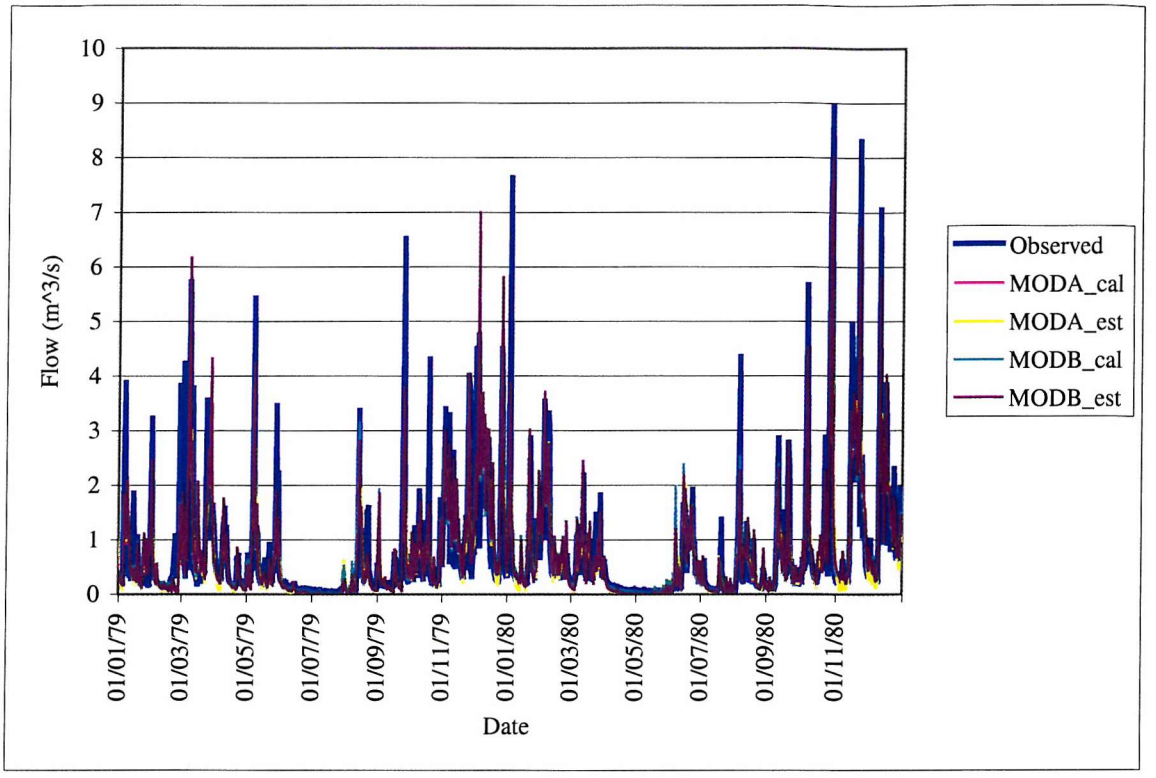



**Figure 8.7 Example observed and simulated hydrographs for 67010.**

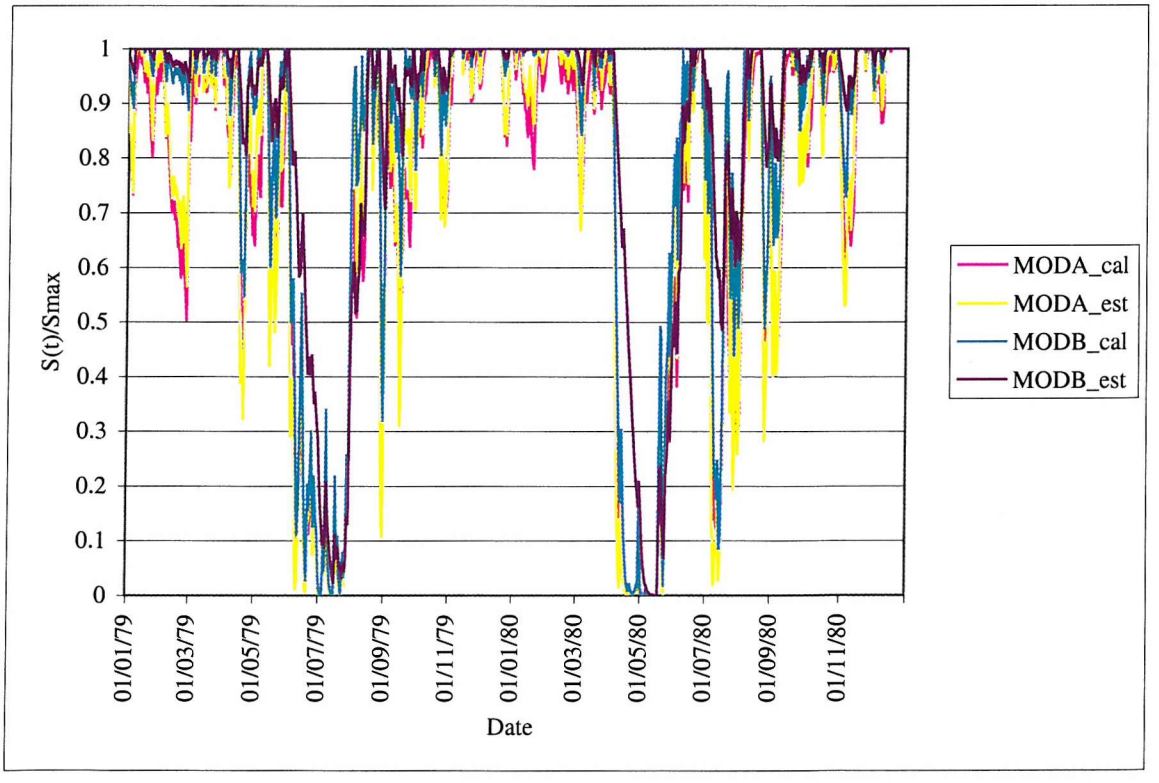



**Figure 8.8 Example graphs of simulated soil moisture status for 67010.**

For MODB the regression based estimate of Cmax is higher than the calibrated one by more than a factor of two. This does not unduly impact on the performance of the model. This indicates that the model is insensitive, in this case, to the value of Cmax. This is a consequence of the fact that significant SMD do not occur frequently within this catchment and hence Cmax plays no major role in ensuring a water balance is closed within this catchment.

### 8.2.5 Discussion

Both models generally simulate the observed flows well, over both the calibration and evaluation periods, when run with the calibrated parameters. The two exceptions to this are this simulation of direct runoff for catchment 44006 by MODA and the simulation of high flows within 34003 by MODB. The former can be attributed to a calibration error whilst the latter is probably a structural error, as MODB also cannot simulate the high flows when run with the regionalised model parameters.

The performance of both models is good when run with the regionalised parameters. The fits obtained are generally not as good as those obtained by using the calibrated parameters. A good water balance is closed within two of the catchments using the regionalised parameters and is closed to with 25% within the other two catchments. The time series obtained using the regionalised models are good, and are very comparable to those obtained using the calibrated parameters. The two exceptions to this are catchment 44006 for MODA and catchment 34003 for MODB. In 44006 the regional form of MODA overestimates Q95 flows and in 34003 MODB is unable to simulate high flows well, leading to poor EFF values for both the regionalised and calibrated parameter vectors.

## 8.3 AN ASSESSMENT OF THE PERFORMANCE OF THE REGIONALISED MODELS ACROSS ALL CATCHMENTS USED WITHIN THE STUDY

### 8.3.1 Evaluation of the quality of the fit of the stream flow simulations obtained using the regionalised model parameters

The assessment of the fit obtained using both model configurations with the regression-based parameter estimates was extended to the full catchment data set. The assessment was made over both the period used for calibrating the models and the corresponding evaluation period. The quality of the model fits obtained using the regression based parameter estimates was assessed using the BIAS, $R^2$, EFF and BEQ95 statistics. Comparing them to the fits obtained using the calibrated model parameters (discussed in Chapter 7) assessed the individual catchment fits. Histograms of these fit statistics over the catchment data set were generated for both model formulations, using the regionalised model parameters, for both the calibration and evaluation periods. The histogram ranges for both BEQ95 and EFF were constrained to omit catchments where the BEQ95 value was greater than 300% and/or EFF was negative. A negative value of EFF indicates that the mean of the observed flows represents a better fit to the hydrograph than the simulated data). Catchments omitted from the EFF histograms are summarised in Table 8.2 for both models. The summary information presented consists of the:

- calibrated and regression-based values of Cmax and, in the case of MODA, b;
- EFF values over the calibration and evaluation period obtained using the calibrated model parameters;
- BIAS values over the calibration and evaluation periods obtained using the regionalised model parameters.

The observed Q95 flow expressed as a percentage of mean flows (Q95% MF) and catchment average SAAR statistics for the catchments are also included within the table. Catchments in which the EFF over the evaluation period for the regionalised model parameters was greater than zero are highlighted in bold. The remaining catchments within the table had negative values of EFF over both the calibration and evaluation periods.

The Q95%MF statistic is a useful measure of the responsiveness of the catchment flow regime. Low values are associated with impermeable, quick response catchments that have a low base flow. Conversely, high values are associated with high base flow, permeable catchments. SAAR is presented as a measure of how wet the catchment is.

**Table 8.2 Catchments with negative values of EFF**

| | Calibrated | | Estimated | | EFF (cal) | | BIAS (reg) | | Q95 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Station | Cmax | b | Cmax | b | Cal | Eval | Cal | Eval | %MF | SAAR |
| *MODA* | | | | | | | | | | |
| 15021 | 85 | 0.22 | 54 | 0.52 | 0.82 | 0.80 | 5 | 2 | 9 | 877 |
| 31017 | 150 | 0.13 | 111 | 0.51 | 0.45 | 0.27 | 35 | 30 | 6 | 647 |
| 34002 | 137 | 0.06 | 96 | 0.62 | 0.23 | 0.45 | 37 | 22 | 23 | 610 |
| 39029 | 762 | 0.37 | 539 | 0.40 | 0.53 | 0.51 | 37 | 30 | 59 | 810 |
| 39065 | 350 | 0.02 | 493 | 0.40 | 0.60 | 0.16 | 105 | 74 | 14 | 695 |
| **42014** | **291** | **0.07** | **215** | **0.41** | **0.10** | **0.57** | 44 | 13 | **18** | **837** |
| *MODB* | | | | | | | | | | |
| **28008** | **429** | 1 | **90** | 1 | **-1.43** | **0.57** | **22** | **-8** | **23** | **1020** |
| 31017 | 417 | 1 | 225 | 1 | 0.13 | 0.32 | 31 | 18 | 6 | 647 |
| 34003 | 251 | 1 | 242 | 1 | 0.17 | 0.29 | 1 | 1 | 51 | 669 |
| 39028 | 785 | 1 | 231 | 1 | 0.91 | 0.83 | 56 | 39 | 38 | 786 |
| 39029 | 494 | 1 | 177 | 1 | -0.57 | -0.62 | 36 | 30 | 59 | 810 |

A common theme with all of the catchments within this table is that they are relatively dry. For MODA the calibrated parameters for Cmax and b indicate that the distribution of catchment stores is skewed towards Cmax. For MODB the calibrated value of Cmax is high. In both cases this equates to a large value of Smax. As discussed in the review of the model behaviour (Chapter 6), this indicates that, within the models of these catchments, significant Soil moisture deficits will not occur frequently. The models will therefore evaporating water from the soil store at the potential rate for a lot of the time. Given that significant soil moisture deficits are likely to occur frequently during the summer within these catchment, this indicates that the models are maximising evaporative losses to compensate for either an overestimation of rainfall or under-estimation of evaporation and/or an overestimate of the contributing catchment area.

The MODA regionalised parameter estimates for the Pareto shape parameter, b, are much higher than the calibrated parameter estimates within these catchments. Furthermore, the MODB regionalised parameter estimates for Cmax are much lower than the calibrated

parameter estimates. In both cases, this results in a lower value of Smax and a modification of the rainfall- runoff relationship. This analysis is confirmed by the observation that, with the exception of 15021 for MODA and 34003 for MODB, the models significantly over-estimate mean flow (large positive BIAS) for all catchments when run with the regionalised parameter estimates.

Considering the time series fit for MODA when the model is run with the calibrated parameters, with the exception of catchments 15021 and 42014, the values of EFF for all of the other catchments are lower than 0.5 over either the calibration or evaluation periods, or both. This pattern is repeated for MODB, with four out of the five catchments listed having low values of efficiency. One of these catchments is the case study catchment, 34003. Whilst this has a low model efficiency, the discussion in the previous section demonstrated that this was a consequence of a poor model fit at high flows only.

The catchments that were excluded from the histograms because the simulated values of BEQ95 were greater than 300% when the models were run with the regionalised model parameters are summarised in Table 8.3. In this table results are only presented for a period (calibration or evaluation) if the use of either the calibrated or the regionalised parameters gave a BEQ95 error greater than 300% over the period in question. Where results are not included for a period, both parameter sets gave BEQ95 values of less than 300%. The observed values for Q95%MF are also presented as a measure of the catchment permeability. Those catchments in which the BEQ95 value obtained using the calibrated parameters was greater than 300%, over either the calibration or evaluation period are highlighted in bold for both models.

All of the catchments within the table are very flashy and/or ephemeral in nature with the majority of Q95 values being less than 5% of the mean flow. Rainfall runoff models generally have difficulty in accurately simulating very small flows. As discussed, a BEQ95 error of 300% will not equate to a large volumetric error in these flashy catchments. A further constraint is that this type of conceptual rainfall runoff model cannot model the behaviour of ephemeral streams as the outflow from the routing reservoirs become asymptotic with the time axis.

**Table 8.3 Catchments with BEQ95 errors greater than 300% over either the calibration or evaluation periods of record**

| Station | BEQ95 errors over the Calibration period (%) Calibrated parameters | BEQ95 errors over the Calibration period (%) Regionalised parameters | BEQ95 errors over the evaluation period (%) Calibrated parameters | BEQ95 errors over the evaluation period (%) Regionalised parameters | Observed Q95%MF |
|---|---|---|---|---|---|
| *MODA* | | | | | |
| **6008** | **369** | **564** | 166 | 365 | **2** |
| **21023** | **298** | **418** | **346** | **519** | **3** |
| 31017 | | | 210 | 741 | 6 |
| 31023 | | | 262 | 741 | 0 |
| 33045 | 178 | 318 | | | 11 |
| **36009** | **1161** | **1679** | **658** | **951** | **2** |
| **39042** | **434** | **459** | | | **9** |
| 39054 | 34 | 352 | | | 5 |
| **39065** | **338** | **973** | 102 | 414 | **14** |
| **40027** | **544** | **689** | | | **2** |
| **46818** | **424** | **699** | | | **0** |
| **47005** | **311** | **336** | | | **4** |
| 47008 | | | 214 | 341 | 4 |
| **54025** | **381** | **556** | 147 | 306 | **3** |
| 55028 | 195 | 437 | | | 9 |
| *MODB* | | | | | |
| 21023 | | | 287 | 392 | 3 |
| **31017** | | | **335** | **654** | **6** |
| **31023** | | | **350** | **654** | **0** |
| **36009** | **1283** | **1309** | | | **2** |
| **40027** | **572** | **644** | | | **2** |
| 46818 | 214 | 566 | | | 0 |
| **39042** | **315** | **437** | | | **9** |
| 54025 | 172 | 377 | | | 3 |
| 6008 | 73 | 371 | | | 2 |

In the majority of catchments, in which the use of the regionalised parameters gave BEQ95 errors in excess of 300% for a model, the use of the calibrated model parameters also resulted in errors in excess of 300%. More catchments had errors of greater than 300% over the calibration period than over the evaluation period. This is a consequence of the evaluation period being generally wetter than the calibration period in most catchments, thus resulting in higher observed Q95 values.

The use of MODA results in a BEQ95 error greater than 300% in a greater number of catchments than the use of MODB. This is influenced by the fact that MODB could not close a water balance in some of these catchments. Whether MODB was successfully

applied to the catchments in which MODA appears to give large errors is summarised in Table 8.4 for both the calibration and evaluation periods. Catchments within the columns titled "yes" represent those catchments in which MODB could close a water balance and the regionalised and calibrated parameters both gave BEQ95 errors of less than 300%. This table shows that for the calibration period more of these catchments were adequately modelled using MODB than not. This is not the case for the evaluation period catchment. This evidence suggests that the use of the regionalised parameters with MODB will result in fewer gross errors at the Q95 flow than the use of the regionalised parameters with MODA.

**Table 8.4 Catchments for which MODB could close a water balance and had BEQ95 error <300% but MODA had BEQ95 errors>300%**

| **Calibration Period** | | **Evaluation Period** | |
|---|---|---|---|
| Yes | No | **Yes** | No |
| 33045 | 21023 | 39065 | 54025 |
| **39065** | 33045 | | 6008 |
| **47005** | 39054 | | 47008 |
| 55028 | | | |

The calibration period histograms of the BIAS, $R^2$, EFF and BEQ95 statistics over the catchment data set are presented in Figure 8.9 for MODA and MODB. The equivalent histograms for the evaluation period are presented within Figure 8.10. These histograms are summarised statistically, in terms of the median and non-parametrically derived 68% confidence interval limits, in Table 8.5. This table also presents statistics summarising the differences between the value for a fit statistic obtained within a catchment using the calibrated parameters and the regionalised parameters. These difference statistics are presented for both the calibration and evaluation periods. The difference between a fit statistic is structured so that a negative value indicates that the use of the regionalised parameters is giving a poorer fit than that obtained with the calibrated parameters. A suffix is added to each statistic to differentiate between periods and to differentiate between the statistics and the differences between those statistics. The nomenclature for these suffixes is presented within Table 8.6

Table 8.5 Summary of regionalised model fit statistics

| | Statistic | MODA | | | MODB | | |
|---|---|---|---|---|---|---|---|
| | | median | 68% u.lim | c.i l.lim | median | 68% u.lim | c.i l.lim |
| **Calibration Period** | BIAS_C_EST | 0 | 14 | -7 | 1 | 13 | -7 |
| | R2_C_EST | 0.86 | 0.90 | 0.79 | 0.86 | 0.91 | 0.80 |
| | EFF_C_EST | 0.70 | 0.80 | 0.54 | 0.70 | 0.79 | 0.58 |
| | BEQ95_C_EST | 21 | 129 | -27 | -19 | 76 | -47 |
| **Evaluation Period** | BIAS_V_EST | -3 | 7 | -12 | -2 | 7 | -11 |
| | R2_V_EST | 0.83 | 0.89 | 0.75 | 0.84 | 0.89 | 0.75 |
| | EFF_V_EST | 0.65 | 0.75 | 0.50 | 0.66 | 0.76 | 0.53 |
| | BEQ95_V_EST | 25 | 116 | -13 | -8 | 70 | -38 |
| **Differences Calibration Period** | BIAS_CD_EST | -4 | -1 | -16 | -3 | 0 | -14 |
| | R2_CD_EST | 0.00 | 0.02 | -0.02 | 0.00 | 0.01 | -0.02 |
| | EFF_CD_EST | -0.01 | 0.03 | -0.11 | 0.00 | 0.03 | -0.07 |
| | BEQ95_CD_EST | 7 | 52 | -35 | -3 | 41 | -27 |
| **Differences Evaluation Period** | BIAS_VD_EST | -1 | 4 | -8 | 0 | 3 | -7 |
| | R2_VD_EST | 0.00 | 0.03 | -0.02 | 0.00 | 0.02 | -0.02 |
| | EFF_VD_EST | -0.01 | 0.03 | -0.09 | -0.01 | 0.02 | -0.06 |
| | BEQ95_VD_EST | -8 | 18 | -55 | -8 | 11 | -48 |

Note 1: The differences are the differences between statistics obtained when using the calibrated model parameters and when using the regionalised model parameters
Note 2: Differences are negative where the use of the regionalised model parameters yields a poorer model fit.

Comparing the objective function histograms over the calibration period (Figure 8.9), it can be seen that the fit statistics for the calibration period are, generally, very similar for both model configurations. This is confirmed by the summary statistics. MODA is marginally better (zero median, but slightly larger 68% confidence interval) in terms of BIAS while MODB is marginally better in terms of $R^2$ and EFF (higher values for the 68% lower limit). MODA has a tendency to over estimate Q95 flows while MODB tendency to under estimate Q95. The 68% c.i. for BEQ95 for MODB is 33 percentile points smaller than that for MODA. This pattern is repeated over the evaluation period. The 68% c.i. for BIAS and BEQ95 over the evaluation period are smaller for both models than over the calibration period. The performance with respect to $R^2$ and EFF is marginally worse over the evaluation period than over the calibration period.

**Table 8.6 Nomenclature for statistic suffixes**

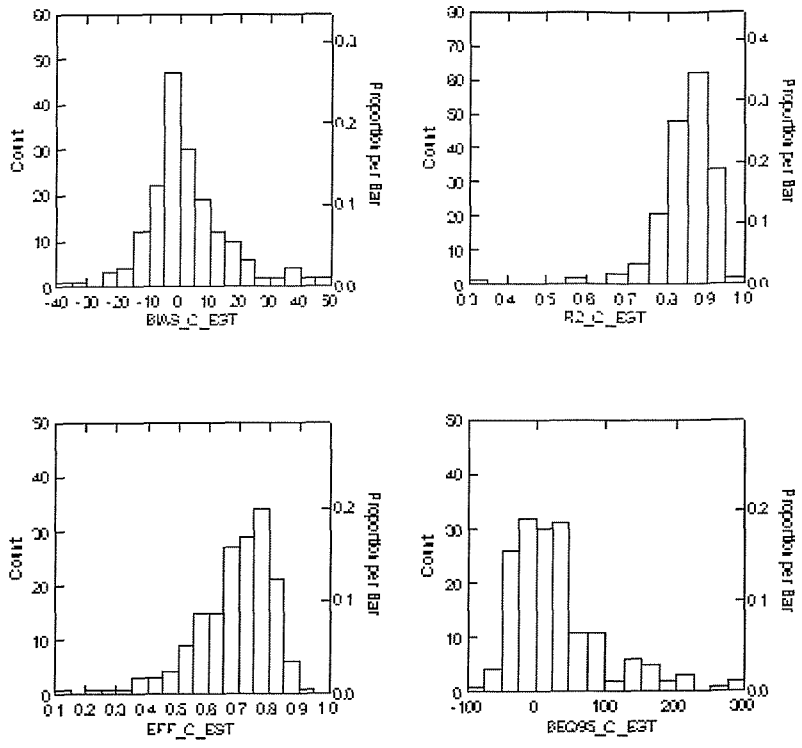
| Suffix | Explanation |
|---|---|
| _C_EST | Value of the statistic obtained using the regionalised model parameters over the calibration period |
| _V_EST | Value of the statistic obtained using the regionalised model parameters over the evaluation period |
| _CD_EST | Difference between the values of the statistics obtained using the calibrated and regionalised model parameters over the calibration period. |
| _VD_EST | Difference between the values of the statistics obtained using the calibrated and regionalised model parameters over the evaluation period. |

Considering the summary statistics in Table 8.5 describing the differences between the fit statistics obtained when using the calibrated and regionalised model parameters, the biggest departures are seen in the two bias statistics: BIAS and BEQ95. The values of BIAS are considerably worse over the calibration period. For MODA, 68% of catchments have an increase in BIAS of between one and 16 percent. This increase is lower for MODB; 68% of catchments have an increase in BIAS of between 0 and –14 percent for MODB. This not surprising as the calibrated model parameters, for the majority of catchments, are constrained to give BIAS values of between ±3% over the calibration period (Chapter 6).

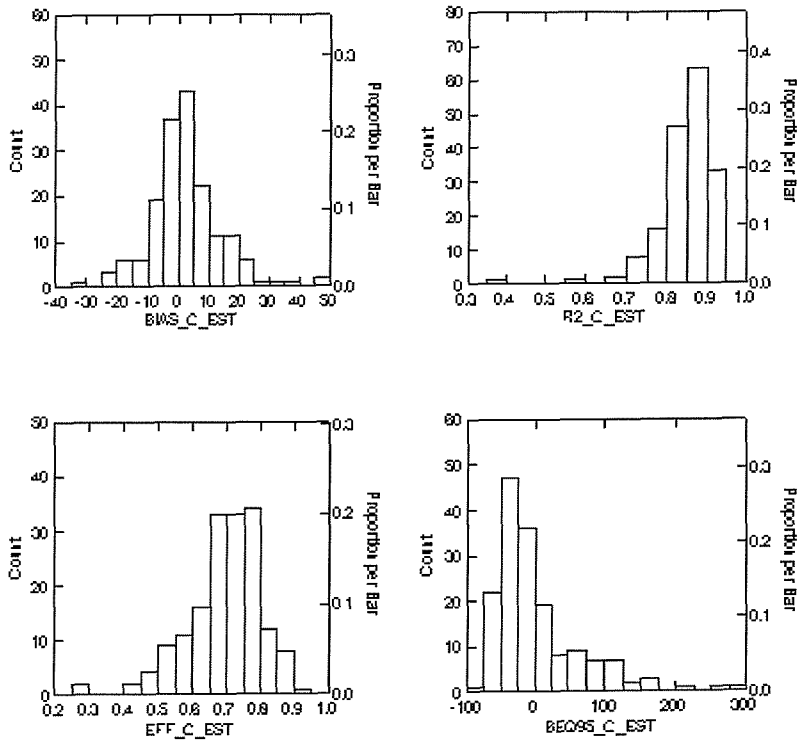
With both models there are a small number of catchments for which the model BIAS is lower when the model is run using the regionalised parameters than with the calibrated parameters. This behaviour is more marked over the evaluation period. Over the evaluation period the calibrated parameters do not necessarily ensure that the models are constrained to close an acceptable water balance. Over the evaluation period the median values for the BIAS_VD_EST statistic are –1 for MODA and zero for MODB. This means that, in 50% of the catchments a larger BIAS is obtained when using the regionalised model parameters and, conversely, a smaller BIAS is obtained in 50% of catchments. Considering the whole catchment data set, this result demonstrates that the distribution of BIAS statistics obtained when the models are run with the regionalised model parameters is comparable to that obtained for the use of the calibrated parameters. However, the asymmetric 68% c.i. implies that, if the use of the regionalised parameters results in a larger value of BIAS then it is likely to be a larger degradation that the improvement observed if the regionalised

parameters give a reduced value for BIAS. This is true for both model configurations.

Following the same logic, the regionalised model parameters give a better fit for BEQ95 over the calibration period than the calibrated model parameters. This is a somewhat surprising result, and is probably a consequence of the trade-off between objective functions during the selection of the calibrated parameter vectors. This does not hold true over the evaluation period, where the regionalised parameters overall result in a poorer fit at Q95 flows. The measure of the time series fit, EFF, and the closely related $R^2$ statistic demonstrate that the time series fit obtained with the regionalised model parameters is very comparable to that obtained with the calibrated parameters, with only a very slight degradation of the fit over 68% of the catchments. In evaluating the time series fit of the model over the catchment data set, it should be remembered that the simulations obtained running the models with the regionalised model parameters resulted in negative values of EFF for a number of catchments. However, as discussed, there is evidence that there were significant errors in the input data for these catchments.

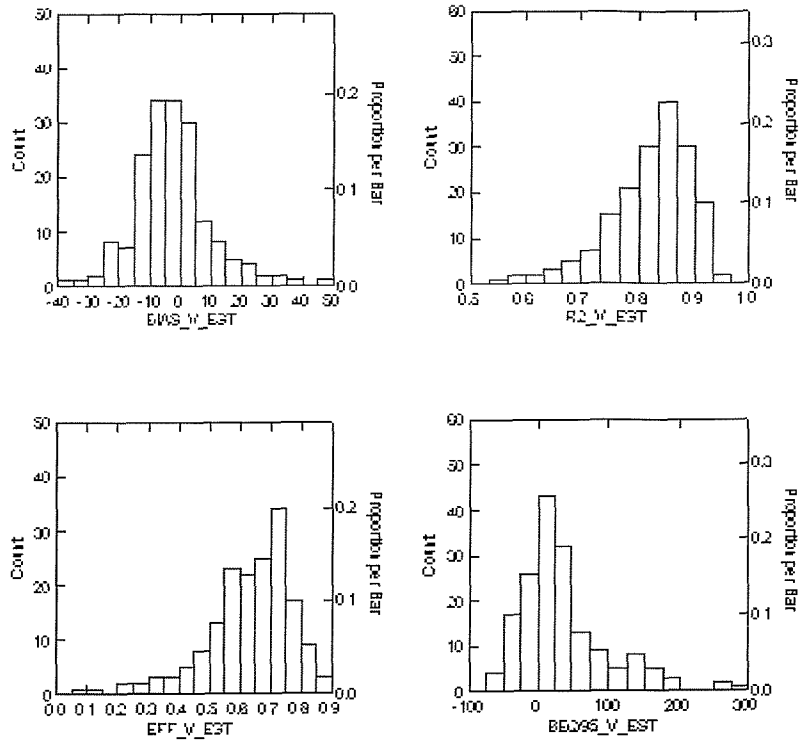



**(a) MODA**

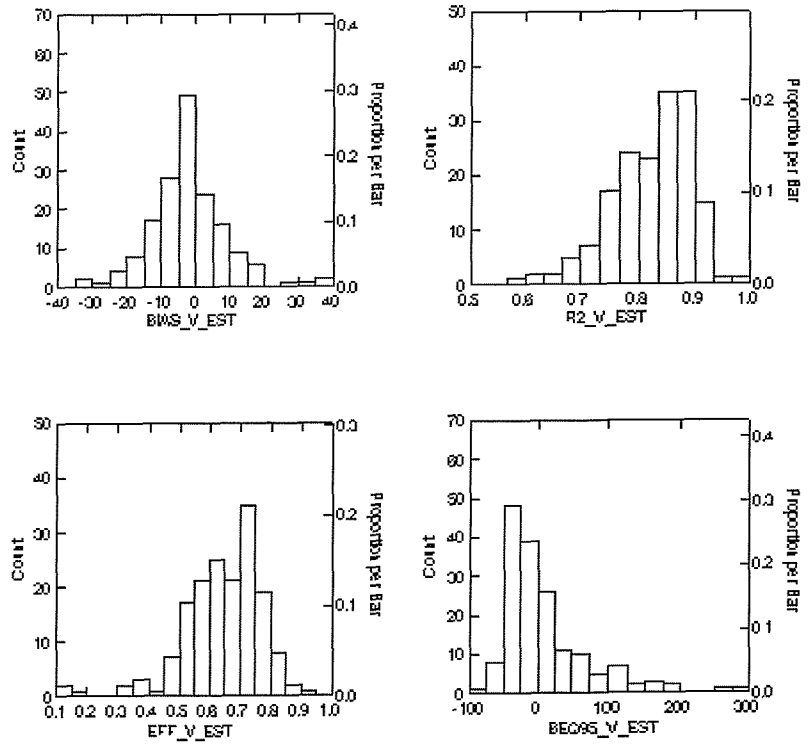



**(b) MODB**

**Figure 8.9 Histograms of calibration period objective functions for the regionalised model parameters.**




**(a) MODA**




**(b) MODB**

**Figure 8.10 Histograms of objective function histograms over the evaluation period for the regionalised parameters.**

### 8.3.2 Relationships between the quality of the regionalised model fits and catchment climate and hydrogeology

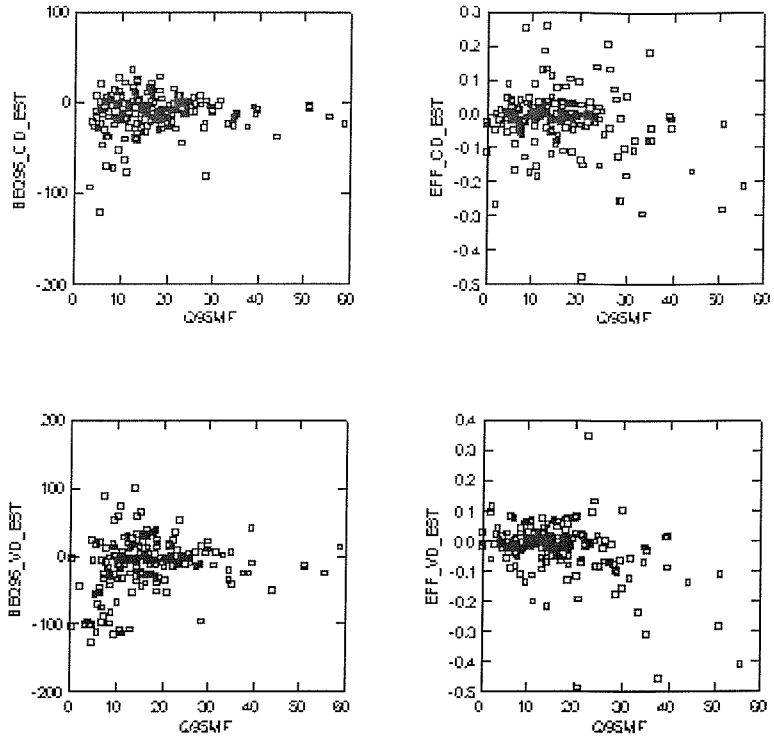
The relationships between the difference in the quality of fit between the calibrated and regionalised model parameters and the climate and hydrogeology of the catchments were investigated. SAAR was used to represent climatic regime and gauged Q95%MF was used as a surrogate variable for hydrogeology. The analysis identified that the magnitude of the differences in BEQ95 and EFF were related to Q95%MF and that the magnitude of the differences in EFF and BIAS were related to how wet the catchments were, as represented by SAAR. The relationship between EFF and SAAR may in part be a reflection of the negative correlation between Q95%MF and SAAR.

The difference statistics for BEQ95 and EFF for individual catchments (summarised statistically in Table 8.5) are graphed in Figure 8.11 as a function of catchment gauged Q95%MF for both model configurations, and over both the calibration and evaluation periods. These plots demonstrate that BEQ95 differences are larger for small values of Q95%MF (flashy catchments) than for permeable catchments over both the calibration and evaluation periods. This is primarily a result of a given numerical error representing a larger percentage difference of the Q95 flows within these flashy catchments, as discussed in previous sections. The departures are generally smaller for MODB than for MODA, indicating the quality of the fit obtained when using the regionalised parameters for MODB is closer to that obtained with the calibrated model parameters across the catchment data set. The distribution of the fit is asymmetric around zero for MODB within catchments with low Q95%MF values, this indicates that negative differences (degradation of model fit) are likely to be larger than positive (improvement of fit) differences. The total variance of departures of MODA is larger than that for MODB and departures are more symmetrically distributed for MODA. For both models there is a greater variation over the evaluation period that for the calibration period.

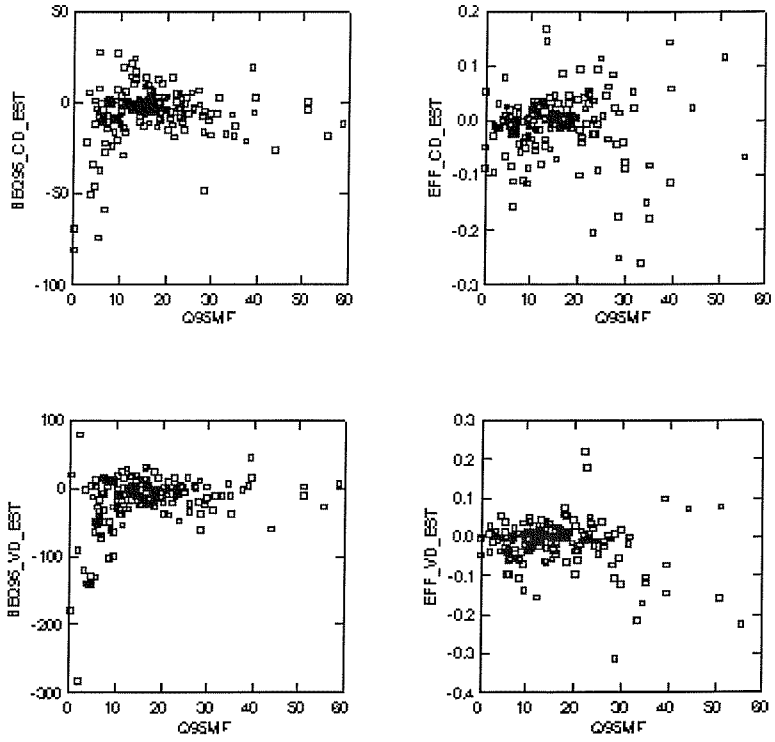
The patterns for the relationships between differences in the EFF statistic and catchment Q95%MF are less well defined. There is a tendency for larger differences to occur in permeable catchments than for impermeable catchments, although it should be noted that departures are both small and relatively symmetrically distributed around zero for both

models. Furthermore, the range of flows within permeable catchments tends to be smaller, and hence the EFF statistic is more sensitive to simulation errors within this type of catchment. Once again, MODA tends to show a greater variation for these differences than MODB and the differences cover a larger range over the evaluation period compared with the range over the calibration period.

The difference statistics for BIAS and EFF for individual catchments (summarised statistically in Table 8.5) are graphed as a function of catchment SAAR within Figure 8.12. These plots are presented for both model configurations, and are presented for both the calibration and evaluation periods. The patterns in the BIAS differences are similar for both models. There is strong relationship with SAAR, with bigger differences observed for catchments with low SAAR values. In these catchments, the models are simulating the complex relationships between soil moisture deficits and evaporation rates and thus the model fits will be sensitive to the values of the soil store model parameters. Over the calibration period the majority of departures are negative - indicating a degradation of fit. The calibrated model parameters constrain BIAS to within 3% for the majority of catchments and to within 5% for a smaller number (particularly for MODA), thus this result is as would be expected. Over the evaluation period the departures are much more symmetrical. Over this period the regionalised parameters perform better than the calibrated parameters in approximately 50% of the catchments, although, as already discussed, the small number of very large departures do tend to be associated with a degradation in fit.

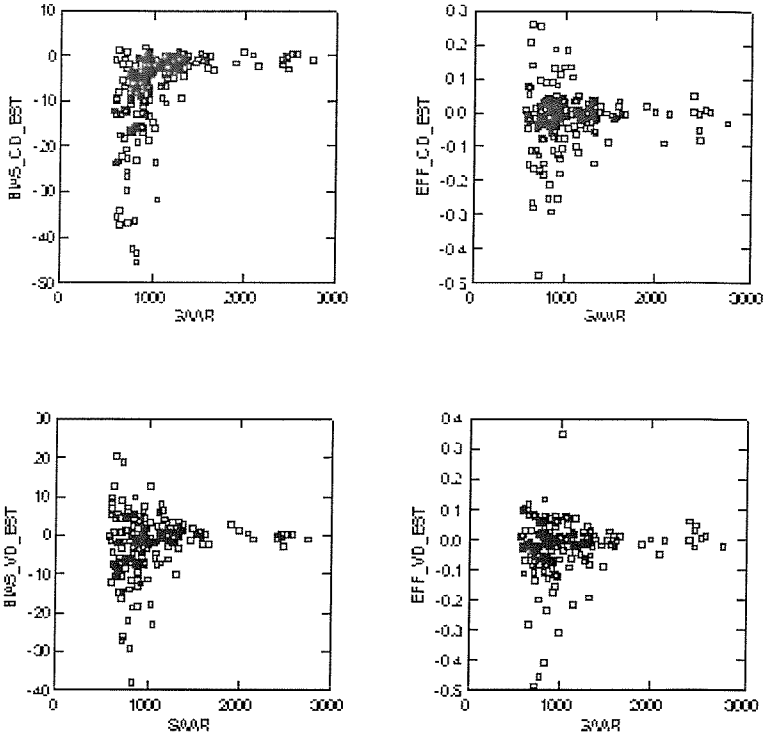



**(a) MODA**

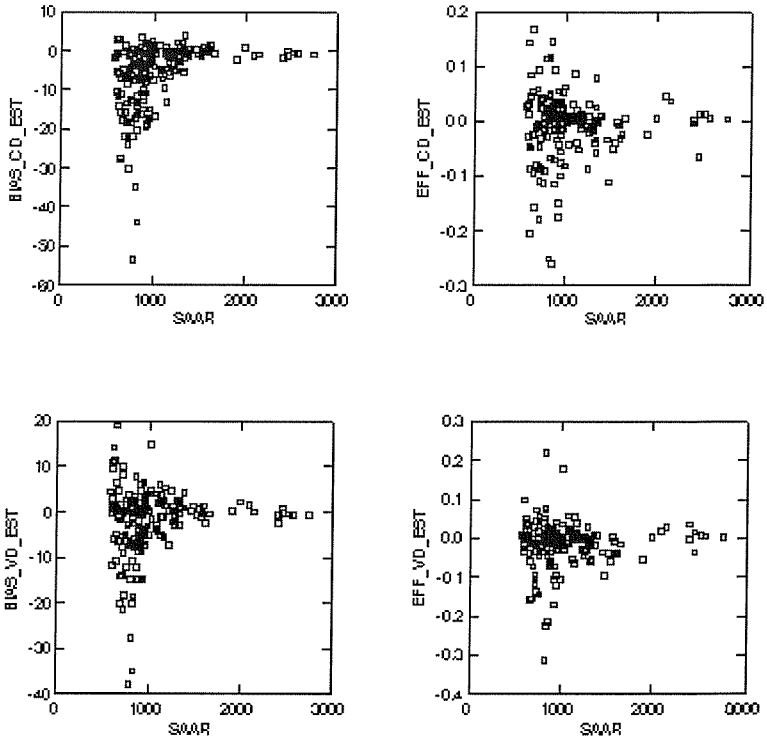



**(b) MODB**

**Figure 8.11 Departures between calibrated and regionalised model values of BEQ95 and EFF plotted as a function of Q95%MF.**




**(a) MODA**




**(b) MODB**

**Figure 8.12 Departures between calibrated and regionalised model values of BIAS and EFF plotted as a function of SAAR.**

For the EFF statistic there is a greater variation in the differences in low runoff catchments. Within the calibration period the distribution of differences is fairly symmetrical for these catchments, with the calibrated parameters giving, on average, better results. This pattern is repeated over the evaluation period except that large departures appear to be restricted to cases where the regionalised parameters give a poorer fit. The differences in EFF tend to be smaller for MODB than MODA, once again indicating that the fit obtained for MODB when using the regionalised parameters tends to be more stable than that for MODA. The value of EFF, as it is a sum of squares based statistic, is sensitive to the fit of the model to large, direct runoff events. The relationship between EFF and SAAR is probably a function of the role that the loss module plays in determining direct runoff corresponding to rainfall events during periods in which there are appreciable soil moisture deficits. The simulations of direct runoff events will be more sensitive to the correct estimation of parameter values for the soil store within these catchments. The picture is further complicated by the fact that the measure is also more sensitive in permeable catchments and that these catchments tend to be located within the lower rainfall areas of the UK.

## 8.4 CLASSIFICATION OF MODEL RESULTS FOR WATER RESOURCES ASSESSMENT

The analysis of results, presented within Sections 8.2 and 8.3, have examined the goodness of fit measures individually. To examine the ability of the models to both close an adequate water balance and maintain a reasonable hydrograph fit when run with the regionalised parameter estimates the catchments were partitioned into classes based on the BIAS and EFF goodness of fit statistics over the evaluation period. The evaluation period was selected as the calibration period was used to calibrate the model parameters that were subsequently used to develop the regionalised relationships for estimating parameters from catchment characteristics. This is not thought to be an important consideration as the regression models for predicting model parameters were rigorously tested for stability, as described in Chapter 7.

A classification of catchments is presented in Table 8.7. This table presents the class limits, the numbers of catchments falling within each class for each model, the percentage of the total number of catchments within each class and the cumulative frequency as the class

interval moves from small BIAS, high EFF to high BIAS, low EFF. The frequency statistics facilitate comparisons between the model as eight more catchments were successfully modelled with MODA than MODB. It should also be noted that two catchments, 12008 and 18014, had less than three years of record over the evaluation period thus reducing the number of catchments to 176 for MODA and 168 for MODB. The class intervals were based on four main BIAS classes and one EFF class. The main BIAS classes are:

- Class A: model BIAS of less than 5%, equivalent to a good quality gauged flow record;
- Class B: model BIAS between 5 and 10%, the 10% level is equivalent to the acceptable artificial influence allowed during catchment selection (Chapter 5);
- Class C: model BIAS between 10-20%, approximately equivalent to the accuracy of a Micro LOW FLOWS estimate of mean flow;
- Class D: model BIAS greater than 20%, classed as a large BIAS error.

Micro LOW FLOWS is an software package used by regulators within the UK for estimating natural and artificially influenced flow statistics within ungauged catchments (Young *et al*, 2000). Each of the first three BIAS classes was subdivided into two EFF classes; a good fit class (Class 1) in which the EFF is greater than 0.6 and an acceptable fit class (Class 2) in which the EFF was between 0.5 and 0.6. One additional class, Class E, was used for catchments in which the model failed to explain more than 50% of the variance (EFF<0.5), irrespective of the value of BIAS. The threshold of EFF<0.5 is conservative. Many catchments with EFF values of between 0.3 and 0.5 had low BIAS errors, high $R^2$ values and good stream flow fits during recession periods. These model fits would be perfectly acceptable for many water resources applications. An example of this is the case study catchment 34003 as discussed within Section 8.2.

Table 8.7 demonstrates that MODB outperforms MODA with respect to the numbers of catchments lying within Class A and Class B. There are slightly fewer catchments in Class C for MODB than MODA. There are 17 catchments that cannot adequately close a water balance for MODA (Class D) compared with seven for MODB. All seven catchments within Class D for MODB are also in Class D for MODA. Of the remaining 10 MODA catchments, for which MODA could not close an adequate water balance, only two were

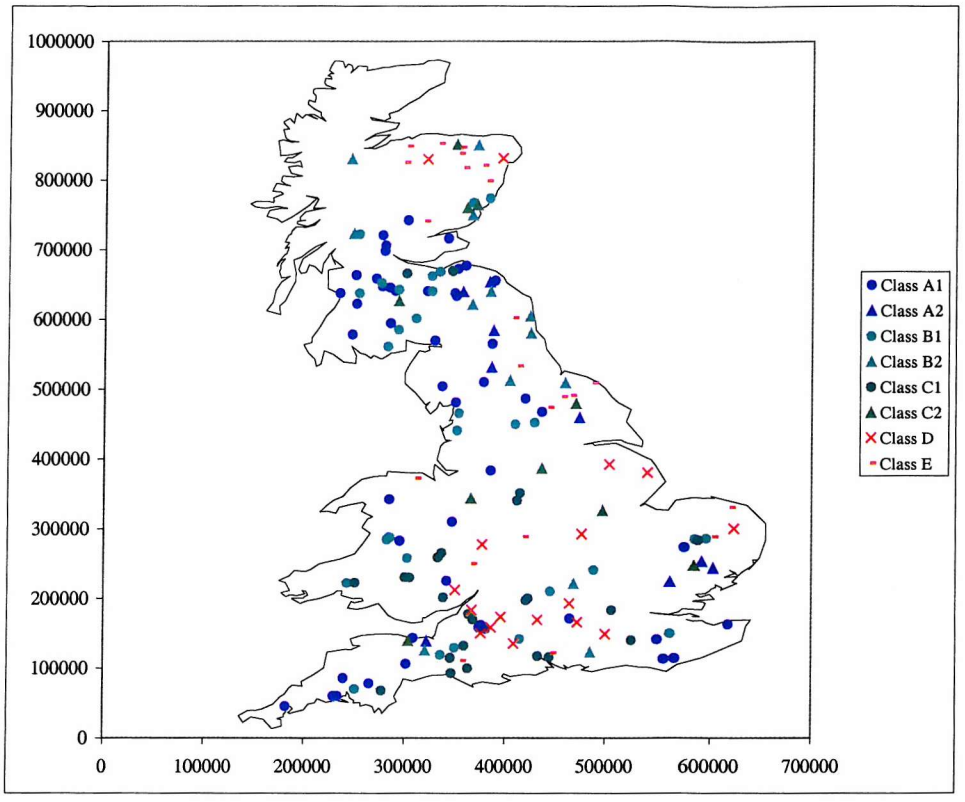
excluded from the catchment set for MODB as a result of the model being unable to adequately close a water balance during calibration. The number of catchments failing to give EFF values greater than 0.5 (Class E) was broadly similar for the two model configurations. The value of EFF was between 0.3 and 0.5 in 17 of the 19 catchments within Class E for MODA. For MODB the model EFF was between 0.3 and 0.5 for 13 out of the 19 catchments within Class E. The number of catchments that had completely unacceptable time series fits will therefore be considerably lower for both models than the membership of class E suggests.

The spatial distribution of the catchments, by class, is presented in Figure 8.13 for both models. This demonstrates that Class D and Class E model fits obtained with MODA tend to be associated with dry catchments on the east of the country (primarily low EFF values) and in the south (primarily high BIAS values). A satisfactory model fit (classes A, B or C) is obtained with MODB for many of the catchments classed as D (large BIAS) for MODA. The catchments for which MODB does not close an adequate water balance tend to be unconfined chalk catchments. The implicit assumption of a closed water balance is likely to be violated within these catchments.

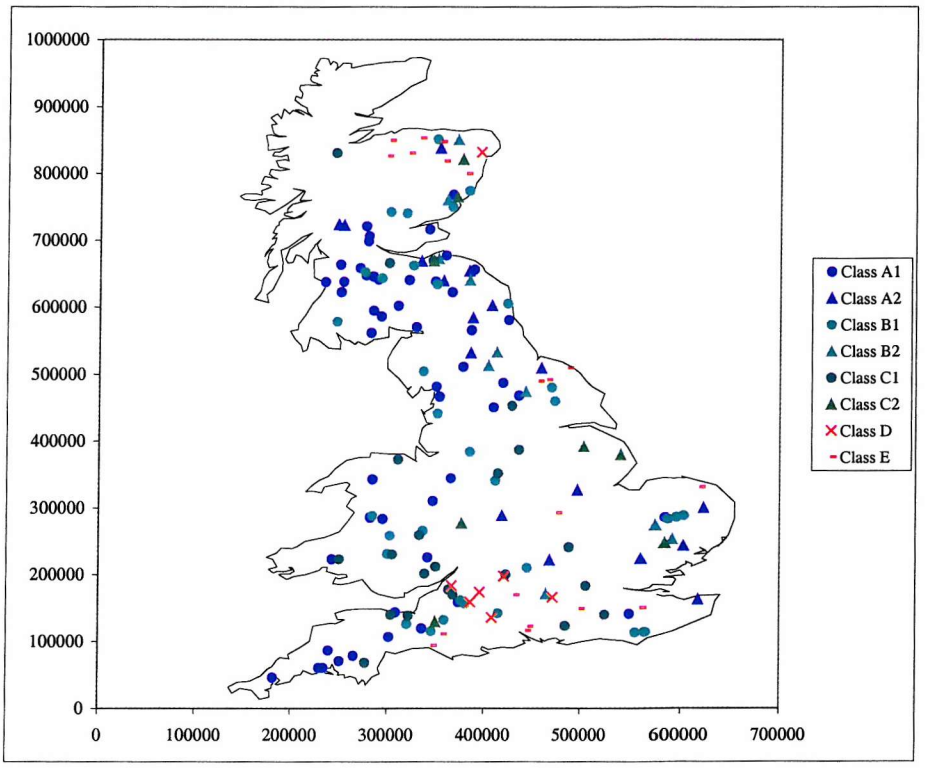
There is a cluster of catchments in the north east of Scotland, which have Class E value of EFF. On closer inspection, the rain gauge networks in these catchments was found to be very sparse raising a question about the accurate spatial representation of precipitation events leading to direct runoff. This in turn could lead to the poor modelling of higher flow events, and hence poor efficiencies.

**Table 8.7 Model fit classification based on evaluation period BIAS and EFF statistics**

| Class | BIAS | EFF | No. in class | % in class | cum.freq (%) |
|---|---|---|---|---|---|
| | ***MOD A*** | ***n=176*** | | | |
| **A1** | <5 | >0.6 | 50 | 28 | 28 |
| **A2** | | 0.5-0.6 | 9 | 5 | 34 |
| **B1** | 5-10 | >0.6 | 30 | 17 | 51 |
| **B2** | | 0.5-0.6 | 13 | 7 | 58 |
| **C1** | 10-20 | >0.6 | 25 | 14 | 72 |
| **C2** | | 0.5-0.6 | 10 | 6 | 78 |
| **D** | >20 | | 17 | 10 | 88 |
| **E** | | <0.5 | 22 | 13 | 100 |
| | ***MOD B*** | ***n=168*** | | | |
| **A** | <5 | >0.6 | 52 | 31 | 31 |
| | | 0.5-0.6 | 17 | 10 | 41 |
| **B** | 5-10 | >0.6 | 33 | 20 | 61 |
| | | 0.5-0.6 | 10 | 6 | 67 |
| **C** | 10-20 | >0.6 | 22 | 13 | 80 |
| | | 0.5-0.6 | 8 | 5 | 85 |
| **D** | >20 | | 7 | 4 | 89 |
| **E** | | <0.5 | 19 | 11 | 100 |




**(a) MODA**




**(b)MODB**

**Figure 8.13 The spatial distribution of model fit classes across the UK.**

## 8.5 PERFORMANCE OF REGIONALISED MODELS ON TWO CASE STUDY ANGLIAN CATCHMENTS

The analysis of the fit of the regionalised rainfall runoff models within the preceding sections has focussed upon the catchments used to develop the regression models for predicting the regionalised parameters. This analysis is certainly valid as the split sample testing of the regression equations for individual parameters demonstrated that the equations were robust when applied to an independent sample set.

As an independent test, the regionalised model forms for MODA and MODB were applied to two of the five case study Anglian catchments used in the model evaluation work presented in Chapter 3. As these catchments were not used in the regionalisation study they represent a totally independent test data set. Furthermore, the results obtained using the models evaluated within the catchments provide a useful set of results against which to compare the performance of the two regionalised models within these catchments. This is a stringent test of the regionalised models as the evaluation models were specifically calibrated for use within these catchments.

The five catchments are fully described in Chapter 3. The Babingley Brook and Sapiston Brook catchments were selected for a comparison of simulations obtained using the regionalised model forms with those obtained using the four models (PDM, HYSIM, IHACRES and TCM) that were calibrated and evaluated within these catchments. The Box catchment was not considered for the analysis, as it was not possible to obtain adequate simulations within this catchment using any of the models evaluated within Chapter 3. The Blackwater and Nene catchments were excluded due to the degree of artificial influences within the catchments and questions that have been subsequently raised about the quality of the naturalised flow time series by Environment Agency staff (Cadman, *pers comms*).

The regionalised rainfall runoff models, MODA and MODB, were applied to the two catchments using rainfall and PE data derived using the methods described in Chapter 2. The rainfall inputs were therefore the same as those used for the model evaluation work but the PE inputs were derived using monthly MORECS data interpolated to a 1km grid. The main impact of these differences was in the Sapiston catchment. The PE estimates derived using

the interpolated MORECS data gave annual PE estimates that were 20% higher than the weekly MORECS data used for the evaluation work within Chapter 3.

The regionalised models were applied over both the calibration and evaluation periods used in Chapter 3 and the model fits assessed using the BIAS, $R^2$, EFF and BEQ95 statistics. Values for the first three of these were available from the results obtained for the application of the PDM, IHACRES, HYSIM and TCM models. These statistics are summarised for the Babingley Brook and the Sapiston Brook within Table 8.8 and Table 8.9 respectively. Results are presented for both the calibration and evaluation periods. The figures within the tables demonstrate that the fit obtained with MODA and MODB over the evaluation period is very comparable to that obtained with the PDM within the Babingley catchment, better that that obtained for the TCM but not as good as that obtained with IHACRES and HYSIM. Within the Sapiston the fit over the evaluation period is again comparable to that obtained with the PDM but better that that obtained with the other models.

**Table 8.8 Goodness of fit statistics for the Babingley Brook**

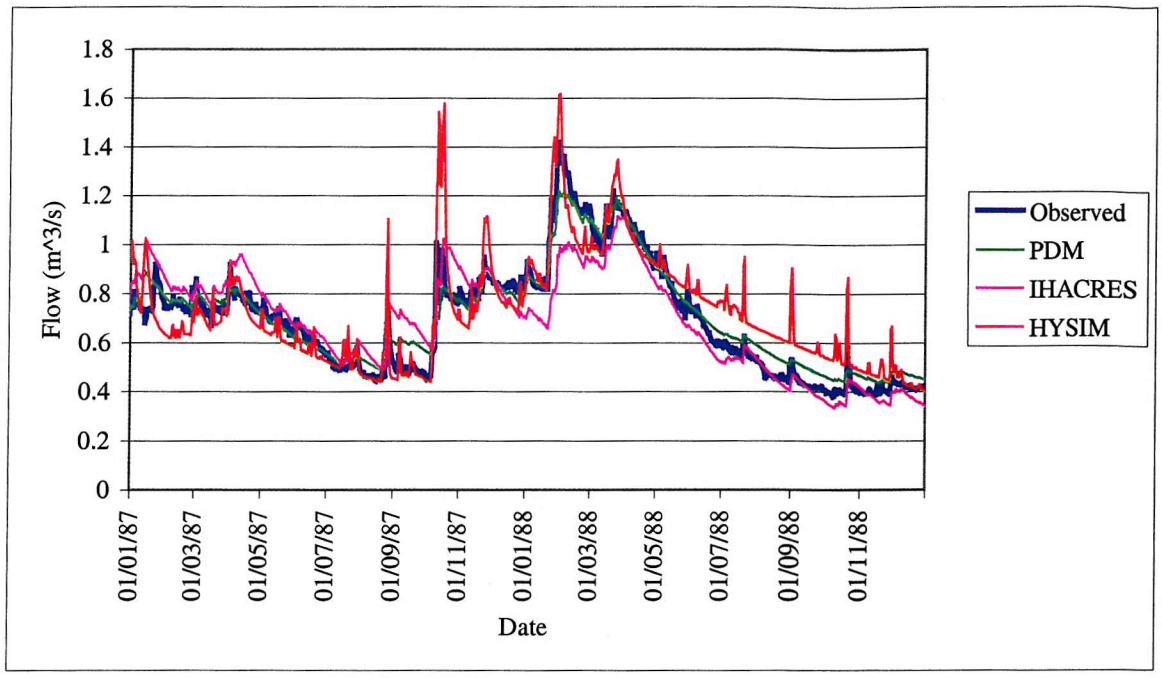
| | PDM | IHACRES | TCM | HYSIM | MODA | MODB |
|---|---|---|---|---|---|---|
| **Calibration** | | | | | | |
| BIAS | 1.97 | -2.73 | 15.61 | 0.76 | -1.7 | -3.37 |
| $R^2$ | 0.95 | 0.88 | 0.88 | 0.86 | 0.91 | 0.95 |
| EFF | 0.87 | 0.72 | 0.67 | 0.73 | 0.72 | 0.41 |
| BEQ95 | | | | | -19.1 | 8.08 |
| **Evaluation** | | | | | | |
| BIAS | 2.11 | -1.58 | 24.91 | 2.46 | -3.86 | -9.76 |
| $R^2$ | 0.93 | 0.89 | 0.66 | 0.88 | 0.85 | 0.89 |
| EFF | 0.68 | 0.78 | 0.27 | 0.75 | 0.66 | 0.51 |
| BEQ95 | | | | | 9.41 | -12.99 |

Over the calibration period, the fit obtained for MODA in the Babingley Brook is not as good as that obtained with the PDM but is similar to that obtained with IHACRES and HYSIM, and is better than that obtained with the TCM. MODB closes the water balance over the calibration period but the time series fit is not as good as that obtained with the other models, although the fit at low flows is better than that obtained with MODA. In the Sapiston catchment, MODA and MODB significantly under estimate mean flow over the calibration period (negative BIAS) by more than 10%. The time series fit for MODA is comparable to that obtained with HYSIM, but is not as good as that obtained with the PDM.

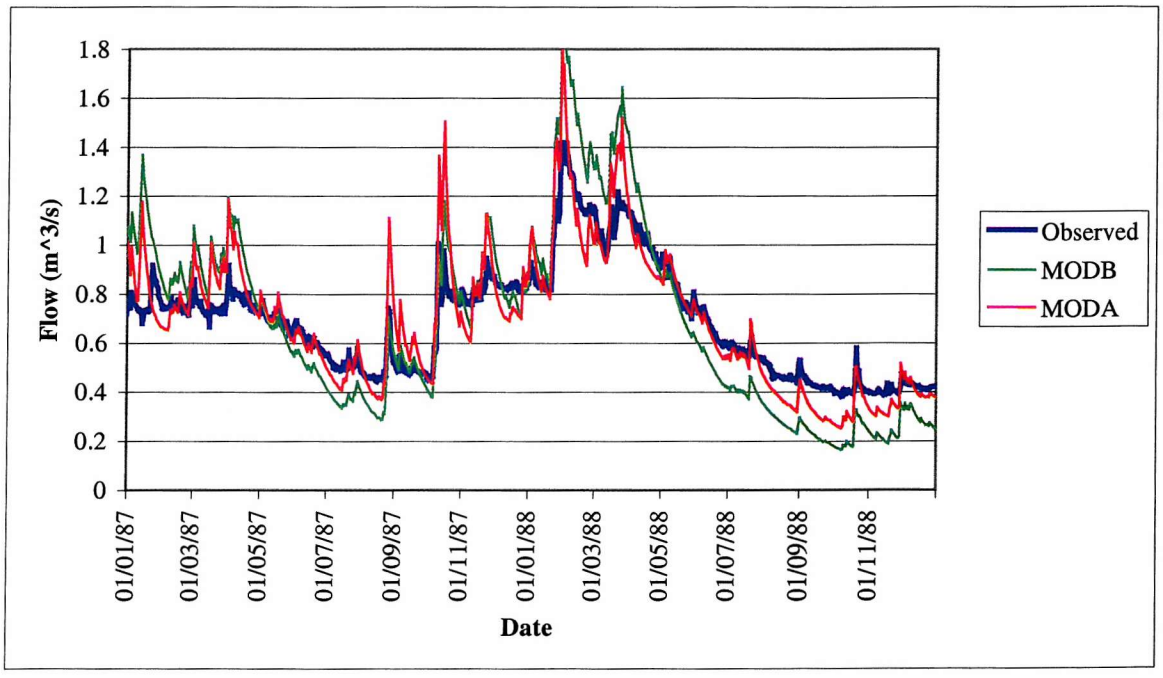
**Table 8.9 Goodness of fit statistics for the Sapiston Brook**

| | PDM | IHACRES | TCM | HYSIM | MODA | MODB |
|---|---|---|---|---|---|---|
| **Calibration** | | | | | | |
| BIAS | -7.49 | 4.16 | 2.50 | 5.10 | -11.25 | -12.7 |
| $R^2$ | 0.91 | 0.72 | 0.25 | 0.85 | 0.85 | 0.82 |
| EFF | 0.77 | 0.27 | -0.30 | 0.67 | 0.67 | 0.59 |
| BEQ95 | | | | | 22.62 | 29.82 |
| **Evaluation** | | | | | | |
| BIAS | -18.94 | 6.90 | 28.28 | 11.50 | 0.43 | -9.35 |
| $R^2$ | 0.84 | 0.71 | 0.16 | 0.82 | 0.8 | 0.79 |
| EFF | 0.65 | 0.17 | -1.49 | 0.55 | 0.64 | 0.61 |
| BEQ95 | | | | | 113.51 | 56.78 |

Example hydrographs from the years 1987 and 1988 for the Babingley Brook are presented in Figure 8.14a for the PDM, HYSIM and IHACRES models and in Figure 8.14b for MODA and MODB. The TCM is not presented, as the model fits obtained with the TCM were poor. It is important to note that this period was part of the calibration period used within the model evaluation studies, and thus represents a calibrated fit for the three evaluation models. The fit of the evaluation models is discussed in detail within Chapter 3, however the figure demonstrates that, visually, the best hydrograph fits are obtained with PDM and IHACRES. With all models the catchment is essentially modelled as a first order catchment, with most of the effective rainfall being routed through a slow flow routing path (the catchment is extremely permeable). The hydrograph fits obtained for MODA and MODB demonstrate that the estimated time constant for the slow flow routing reservoir is too low, resulting in steeper recessions rates and quicker response to recharge than that observed. The same hydrographs are presented for the Sapiston Brook in Figure 8.15a and Figure 8.15b. In this catchment the best calibrated model fit is for the PDM, IHACRES fails to simulate the base flow and the recession rates obtained with HYSIM are too low. Furthermore, all models tend to under predict the highest observed flows. The hydrographs obtained with MODA and MODB are both excellent, from a water resources perspective, with the models simulating all the observed features of the hydrograph except for a tendency to significantly under predict high flow events. This latter behaviour is the reason for the under prediction of mean flow by MODA and MODB over this calibration period.

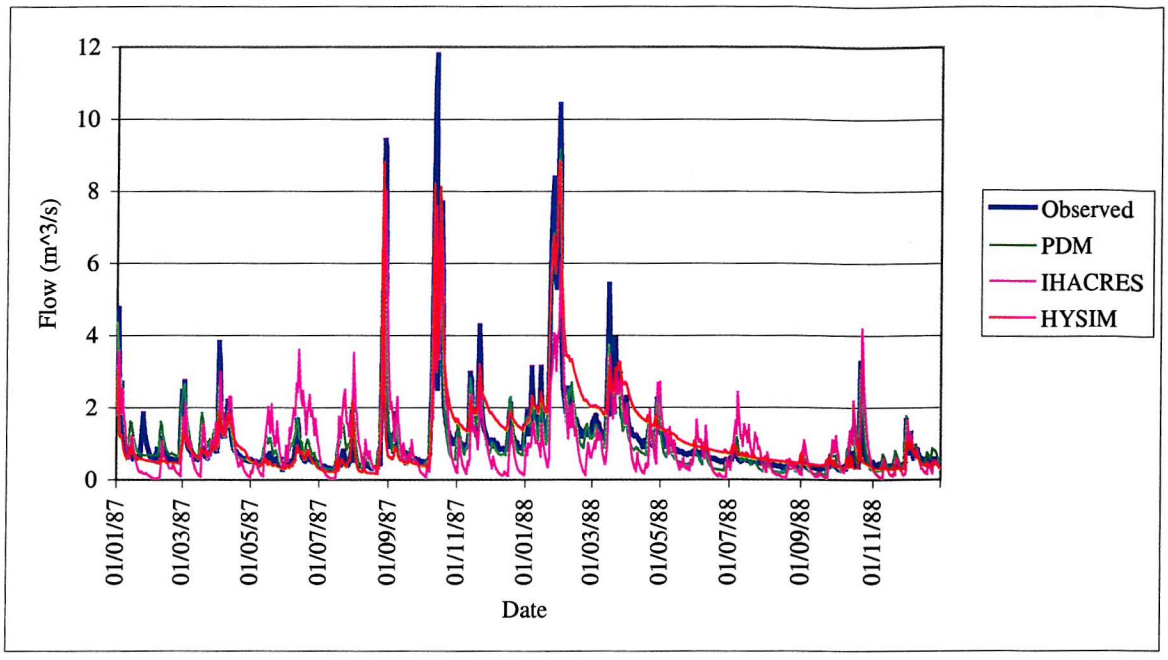



**(a)**

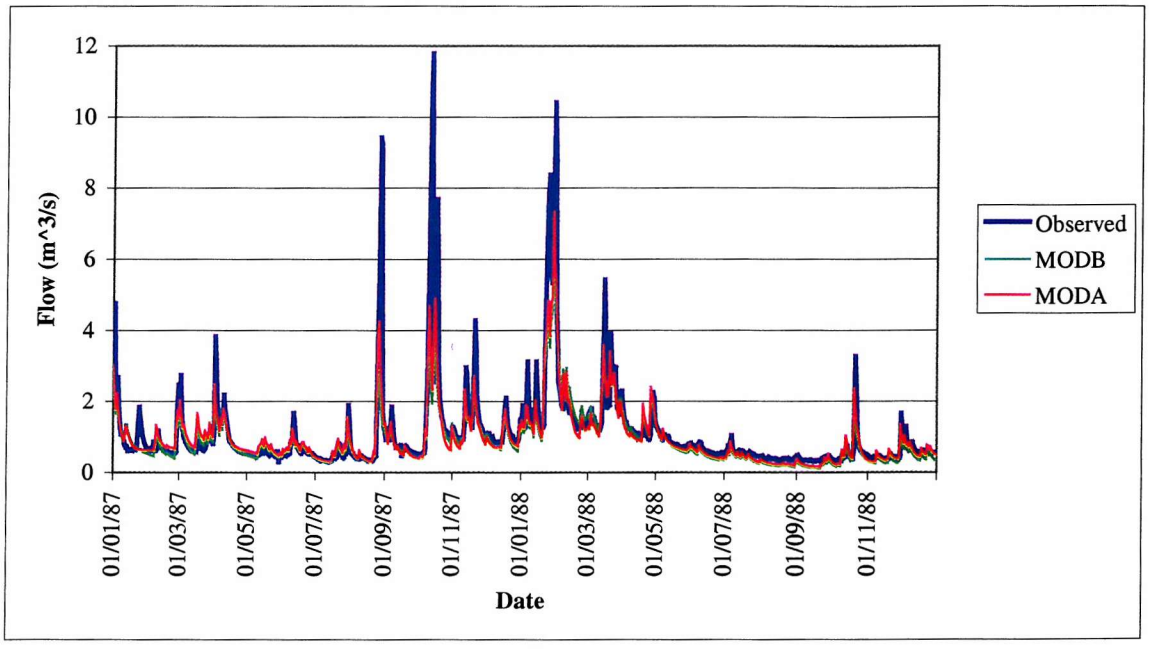



**(b)**

**Figure 8.14 Example observed and simulated hydrographs for the Babingley Brook.**




**(a)**




**(b)**

**Figure 8.15 Example observed and simulated hydrographs for the Sapiston Brook.**

## 8.6 SUMMARY

The objective of the analysis presented within this Chapter was to assess the performance of the MODA and MOB models when run with parameters predicted using the regression based models described within Chapter 7. The performance of the models was assessed by:

1. evaluating the fits obtained within the case study catchments, considered in Chapter 6, with respect to the observed flows and the simulated flows obtained with the calibrated model parameters;
2. evaluating general patterns in the quality of model fit obtained using the regression based parameter estimates across the catchments used to originally calibrate the models;
3. developing a classification scheme for classifying model simulations in the context of their applicability to water resource assessments;
4. comparing the performance of the regionalised model fits within two of the Anglian catchments used in the model evaluation studies presented in Chapter 3.

These assessments are discussed in detail within the preceding sections, and are summarised here. The objective of the first pair of assessments was to compare the performance of the regionalised rainfall runoff models with what could be expected from the same model structures if stream flow data were available for calibration purposes. The objective of the second pair was to assess whether the errors in the simulated stream flows, obtained using the regionalised models, are sufficiently small for the techniques to be a useful aid in the management of water resources within the UK. The models that were evaluated within the Anglian case study catchments are all in general use with practitioners within the field.

Considering the behaviour of the models within the case study catchments (Section 8.2), both models generally simulated the observed flows well over both the calibration and evaluation periods when run with the calibrated model parameters. The exception to this was the performance of MODB within catchment 34003 where the calibrated model failed to simulate high flows well leading to a poor value of efficiency. The fits obtained for both models, by running the models with the regionalised parameters were generally not as good

as those obtained by using the calibrated parameters. A good water balance was closed within two of the catchments using the regionalised parameters and was closed to within 25% within the other two catchments. With two exceptions (catchment 44006 for MODA and 34003 for MODB), the time series obtained using the regionalised models was good and were very comparable to those obtained using the calibrated parameters.

The assessment of the fit obtained using both model configurations with the regression-based parameter estimates was extended to the full catchment data set within Sections 8.3 and 8.4. For each catchment, the assessment considered both the periods used for calibrating the models and the corresponding evaluation periods. The quality of the model fits obtained using the regression based parameter estimates was assessed using the BIAS, $R^2$, EFF and BEQ95 statistics. The assessment considered:

- histogram analysis of the variation in these fit statistics over the full catchment data set over both the calibration and the evaluation periods;
- a comparison of the quality of the regionalised model fits obtained within individual catchments with the fits obtained using the calibrated model parameters (discussed in Chapter 7);
- the relationships between the climate and hydrogeology of the study catchments and the differences in model fit observed between the regionalised and calibrated models;
- classification of the suitability of the regionalised models for support water resource assessments within the UK. This classification considered both model BIAS and EFF statistics.

Examples of catchments for which simulations had negative values of EFF were observed for both MODA and MODB. A common theme with all of the catchments, where this occurred, is that they were relatively dry catchments. From a knowledge of model behaviour, there was evidence of the calibrated model parameter vectors, for both models, compensating for errors in climate (overestimation of rainfall or under-estimation of evaporation) and/or an overestimate of catchment area to ensure the model could close a water balance. The regionalised model parameters therefore differed significantly from the calibrated parameters within these catchments, leading to large errors in the estimation of effective rainfall and hence the water balance. These errors propagate through the model

resulting in the poor values observed for the EFF statistic.

Errors in BEQ95 greater than 300% were identified for a number of catchments for both regionalised models. All of the catchments were very flashy and/or ephemeral in nature. A BEQ95 error of 300% does not equate to a large volumetric error in these flashy catchments. Rainfall runoff models generally have difficulty in accurately simulating very small flows; the constraint is that the outflow from the routing reservoirs becomes asymptotic with the time axis and thus a zero outflow occurs only at infinity. In the majority of catchments, the use of the calibrated model parameters also resulted in errors in excess of 300%. There was evidence that the regionalised form of MODB gave fewer gross errors at Q95 than MODA in these flashy catchments.

Considering the variation in fit statistics over the calibration period, the fit of the two regionalised models was judged to be very similar. MODA is marginally better in terms of BIAS while MODB is marginally better in terms of the overall time series fit and significantly better than MODA in simulating low flows. Over the evaluation period both models generally have better fits for BIAS and BEQ95, indicating that they are better at predicting mean flow and low flows over the evaluation period. However the overall time series fit was marginally worse.

Comparing the fit of the regionalised models with the calibrated models identified that over the calibration period the calibrated parameters resulted in much lower water balance errors. This was as would be expected, as closing a water balance was a key part of the calibration scheme. Over the evaluation period the BIAS statistics obtained with the regionalised models were very comparable to those obtained for the calibrated model parameters. There was evidence for both model forms that, if the use of the regionalised parameters resulted in a poorer BIAS then it was likely to be much poorer than the improvement observed if the regionalised parameters give a reduced value for BIAS.

The regionalised model parameters give a better fit for BEQ95 over the calibration period than the calibrated model parameters. This is a somewhat surprising result and is probably a consequence of the trade-off between objective functions during the selection of the calibrated parameter vectors. Over the evaluation period, the use of the regionalised

parameters overall tended to give a marginally poorer fit at Q95 flows that the fit obtained using the calibrated model parameters. The overall time series fit obtained with the regionalised model parameters was found to be, generally, very comparable to that obtained with the calibrated parameters over both periods, with only a very slight degradation of the fit over 68% of the catchments.

The differences (corresponding to both better and worse simulations) between the values of the BEQ95 statistic obtained with the regionalised and calibrated parameters tend to be larger for impermeable catchments over both the calibration and evaluation periods. This is primarily a result of a given numerical error representing a larger percentage difference of the Q95 flows within these flashy catchments, as previously discussed. The departures are generally smaller for MODB than for MODA indicating the quality of the fit obtained when using the regionalised parameters for MODB is closer to that obtained with the calibrated model parameters across the catchment data set. For both models there is a greater variation in these differences over the evaluation period that for the calibration period.

There is a tendency for larger differences in efficiency to occur in permeable catchments than for impermeable catchments, although it should be noted that the differences were small. The differences for MODA tended to be larger than for MODB and the differences covered a larger range over the evaluation period compared with the calibration period.

The analysis also identified relationships between catchment SAAR and the difference statistics for BIAS and EFF for individual catchments. The patterns in the BIAS differences are similar for both models. There was a strong relationship between the magnitude of the BIAS differences with SAAR. Larger differences occur for catchments with low SAAR values. In these catchments the models are simulating the complex relationships between soil moisture deficits and evaporation rates and thus the model fits will be sensitive to the values of the soil store model parameters. Over the calibration period the majority of difference were biased towards a smaller BIAS error being obtained for the use of the calibrated parameters. This is because the calibrated model parameters constrain BIAS to within 3% for the majority of catchments over this period. Over the evaluation period the differences were much more random, with the regionalised parameters giving lower BIAS

statistics than the calibrated parameters in approximately 50% of the catchments.

For the EFF statistic there was a greater variation in the magnitude of differences in low runoff catchments for both models and both periods. The differences in EFF were smaller for MODB than MODA. This gives added weight to the evidence that the fits obtained for the regionalised form of MODB were more stable than those obtained for the regionalised form of MODA. The relationship between EFF and SAAR was attributed to the role that the loss module plays in determining direct runoff corresponding to rainfall events during periods in which there is an appreciable soil moisture deficit. The simulations of direct runoff events will be more sensitive to the parameter values for loss module within these catchments. The picture is further complicated by the fact that the measure is also more sensitive to simulation errors in permeable catchments, which tend to be located within lower rainfall areas of the UK.

The classification of the ability of the models to both close an adequate water balance and maintain a reasonable hydrograph demonstrated that MODB gave acceptable simulations in more of the catchments than MODA. For MODB 67% of catchments were within classes A and B compared with 58% of catchments for MODA. Classes A and B represent simulations that are good, or very good. For MODB 85% of catchment were Class C or better compared within 78% of catchments for MODA. Two further classes, Class D (poor closure of a water balance) and Class E (lower quality time series, EFF statistic <0.5) accounted for the remaining catchments. The value of EFF was between 0.3 and 0.5 in 17 of the 19 catchments within Class D for MODA. For MODB the value of EFF was between 0.3 and 0.5 for 13 out of the 19 catchments within Class D. Many catchments with EFF values of between 0.3 and 0.5 had low BIAS errors, high $R^2$ values and good stream flow fits during recession periods. These model fits would be perfectly acceptable for many water resources applications. An example of this is the case study catchment 34003. The number of catchments that had completely unacceptable time series fits will be considerably lower for both models than the membership of class D suggests.

The spatial distribution of the catchments by class was mapped. This demonstrated that the Class E and Class D model fits obtained with MODA tend to be associated with dry catchments on the east of the country (primarily low EFF values) and in the south

(primarily high BIAS values). A satisfactory model fit (class A, B or C) was obtained with MODB for many of the catchments classed as E (large BIAS) for MODA. Many of the catchments for which MODB does not close an adequate water balance were unconfined chalk catchments. The implicit assumption of a closed water balance is likely to be violated within these catchments.

As a completely independent test, the regionalised model forms for MODA and MODB were applied to the Babingley Brook and Sapiston catchments. These catchments were used within the model evaluation work that is presented within Chapter 3. The results obtained were compared within the results obtained for the four models (PDM, HYSIM, IHACRES and TCM) calibrated and evaluated within these catchments. These catchments were not used within the regionalisation work and therefore represent a totally independent test of the regionalised models.

These Anglian case study catchments are amongst some of the driest gauged catchments within the United Kingdom. The treatment of evaporation and the correct modelling of actual evaporative losses are primary issues when modelling these catchments. The simulations obtained using MODA and MODB, particularly in the Sapiston, demonstrate that the use of the regionalised parameters can simulate the behaviour of soil moisture well. The performance of the regionalised models is very comparable with the results obtained through the bespoke calibration of a range of models on flow data from the catchment. The fit of the regionalised models is not quite as good over the calibration period used for the bespoke model application, as would be expected, but is very comparable over the evaluation period, and in the Sapiston catchment is actually better than three of the four bespoke model applications.

# 9 Evaluation of alternative approaches to estimating historical time series at ungauged river reaches

The estimation of stream flow for ungauged catchments by transposing gauged stream flow data from an analogue catchment is a widely used technique within the water industry. This chapter presents an evaluation of the fit of transposed to observed data for commonly used techniques, and compares the fit obtained using the best of these techniques with that obtained using a regionalised rainfall runoff model. In addition to the commonly used transposition techniques, another technique based on normalising by predicted average annual runoff has also been evaluated. The objective of the comparison was to determine whether is it possible to predict the conditions under which the use of a regionalised rainfall runoff model will result in a better fit than that obtained by transposing data from the best available analogue catchment, and vice versa. The evaluation of methods for transposing gauged flow data from analogue catchments is presented in Section 9.1 The results obtained using the best of these methods are compared with those obtained using MODB within Section 9.2. An evaluation of whether it is feasible to predict when the use of analogue data will provide a better simulation of stream flow is presented in Section 9.3.

## 9.1 EVALUATION OF METHODS FOR TRANSPOSING GAUGED FLOW DATA FROM ANALOGUE CATCHMENTS

### 9.1.1 Methods for transposing gauged flow data to ungauged catchments

In the United States a number of researchers have published techniques for scaling gauged stream flow from an gauged catchment to an ungauged catchment. These techniques draw from systems engineering, and are all of the form:

$$QX_T = fn\left(\frac{A_T}{A_A}\right)QX_A, \qquad \textbf{(9.1)}$$

where:

$QX_T$ = the flow of interest in the ungauged catchment;

$QX_A$ = the flow of interest in the analogue cacthment;

$A_T$ = the catchment area for the ungauged catchment;

$A_A$ = the catchment area for the analogue catchment;

fn = a scaling constant, or function.

Murdock and Gulliver (1993) present the derivation of a scaling function for the upper mid-western states (Minnesota, North Dakota, South Dakota, Wisconsin and Iowa). The paper also explores a methodology for assessing the uncertainty associated with annual stream flow predictions obtained by using the proposed scaling function. Vogel and Sankarasubramanian (2000) present a technique for scaling annual mean flows within the US, this technique is based on the analysis of 1433 gauged flow records and the definition of scaling functions for 17 broad hydrological/climatological regimes covering the US. Hughes and Smakhtin (1996) present a daily stream flow scaling methodology for infilling gaps in flow records by transposing flow data from an analogue catchment. In this method, the scaling is dependent on the relationships between flow duration curves for concurrent periods of record. This is similar in concept to short record techniques, described by NERC (1980) and Shaw (1988), for use within the UK.

All techniques for transposing flow data from an analogue catchment to an ungauged catchment rely on identifying a suitable analogue catchment, and all make the assumption that the flows within the ungauged and analogue catchments are synchronous; i.e. the flows increase and decrease together. Within the UK the spatial heterogeneity of both climate and catchment hydrogeology is high. Both of these issues are important if the objective is to simulate the daily time series of flows at the ungauged site. This has resulted in the use of more deterministic approaches to scaling within the UK that seek to address this heterogeneity. The UK techniques all attempt to adjust the flow record for the analogue for differences in hydrological scale (differences in rainfall and catchment area) between the analogue and ungauged catchments.

An analogue catchment is usually a catchment that is:

- geographically close to the ungauged catchment, and hence has the same climatic regime;
- hydrogeological similar;
- similar in size;
- either a natural catchment, or a catchment for which there is a naturalised flow record.

Ideally, the analogue catchment should lie upstream or downstream of the ungauged catchment. For this study such an analogue would be termed connected. In the case of a connected analogue, there is a strong serial correlation between the flow measured at the analogue gauge and the flows at the ungauged site as the water that flows past both points has a common component. It is quite common that there is no connected analogue available, and thus a catchment from an adjacent system or tributary for the same system is selected as the analogue for the ungauged catchment. For this case the analogue would be termed unconnected.

If there is a short period of measured flow data at the target site, it is common practice to build these data into the transposition methods through the use of regression based relationships between the flows for the target and the analogue catchments (e.g. Shaw, 1988). This study is concerned with the case where there are no measured flow data for the target catchment. The fits obtained using three methods have been evaluated. The methods tested were:

1. Re-scaling by catchment area. The flow record for the analogue catchment is normalised by the topographic catchment area for the analogue catchment and re-scaled by the catchment area for the ungauged, or target catchment.

2. Re-scaling by catchment area and estimated standard period Average Annual Runoff (AARO(61-90)). The AARO(61-90) estimates were derived using the runoff model, developed by Holmes & Young (2000a) to replace the simple water balance model (Gustard *et al*, 1992) currently used within the Micro LOW FLOWS software package (Young *et al,* 2000) for estimating catchment mean flow for ungauged river reaches. In this method the gauged flows for the analogue catchment are normalised by the

catchment area and the catchment estimate of AARO(61-90). The normalised flows are then re-scaled by the target catchment area and estimated AARO(61-90) for the target catchment. As mean flow can be estimated as the product of catchment area and AARO, this technique is equivalent to re-scaling by estimated 1961-90 mean flow. The approach is a refinement of the common practice of re-scaling by area and SAAR. Re-scaling by SAAR implies a linear relationship between the difference in catchment rainfall and the difference in catchment runoff between the two catchments. This approximation may be justified in high rainfall catchments, but it does not take into account differences in evaporation losses, and the non-linear relationships between these losses, rainfall and soil moisture behaviour. This is particularly important in low rainfall catchments where evaporation limiting soil moisture deficits commonly occur within the summer months.

3. Re-scaling by the estimated flow distribution. In this method the differences in the distribution of flows, as represented by differences in an estimated annual Flow Duration Curve (FDC), are taken into account. An estimated FDC is obtained for both the analogue and target catchments, a correction factor for each percentile point is then obtained by dividing though the flow values for each percentile point for the target catchment by the corresponding percentile point flow value for the analogue catchment. The nearest percentile point is then identified for each flow value within the analogue flow record and the flow value is re-scaled by the corresponding percentile correction factor. An important assumption within this method is that the flows within the catchment pairs are synchronous. This has been demonstrated not to be the case at low flows by Young *et al* (2000) for the timing of the occurrence of low flows.

Method 3 is not a commonly used method due to the computational effort used. Methods for estimating FDC at ungauged sites are presented by Holmes & Young (2000b). These methods for estimating the FDC are used within the replacement system for Micro LOW FLOWS, Low Flows 2000, and are based around a region of influence approach. In this approach ten catchments are selected from a pool of 653 source catchments, based on the similarity of the source catchments to the ungauged catchment. Similarity is measured in terms of the fractional extents of HOST classes found within the ungauged catchment. A weighted Euclidean distance measure is used to evaluate similarity. The flow duration curve (standardised as a percentage of mean flow) for the ungauged site is then estimated by taking

a weighted average of the observed standardised flow duration curves for the selected source catchments. The weight for an individual FDC is based upon the Euclidean distance measure for the catchment. The dimensionless curves are then re-scaled by the catchment estimate of mean flow obtained from the topographic catchment area and the estimate of AARO(61-90).

### 9.1.2 The evaluation of transposition methods

The three methods for transposing analogue data were evaluated by selecting analogue catchments for each of the 179 MODA case study catchments, henceforth called target catchments. These target catchments were selected as the process of calibrating the rainfall runoff models on these catchments identified that there were no major errors in the estimates of input data; rainfall, evaporation demand, catchment area and stream flow data. For each method, the flow data from the selected analogue catchment was transposed to the target catchment and the fit of the transposed flows to the observed flows evaluated for the target catchment. The statistics of BIAS, EFF, R2 and BEQ95, used for evaluating the fit of the regionalised rainfall runoff models (Chapter 8), were used to evaluate the fit of the transposed flow record. To avoid confusion with the use of these statistics within other chapters in other contexts, the statistic are prefixed with "A_" to denote that they relate to the transposition of gauged data from analogue catchments.

For each catchment, the potential analogue catchments were identified from the pool of natural catchments identified as potential candidates for the regionalised rainfall runoff model work. The selection of these catchments is described in Chapter 5. Potential analogues were identified for each of the MODA catchments using the following criteria:

- the distance between the analogue and target catchment had to be less than 50km;
- the difference in BFIHOST had to be less than 0.1.

BFIHOST is used as a surrogate for hydrogeological similarity in this context. The figure of 0.1 is the threshold necessary to identify whether two catchment estimates of BFIHOST are significantly different from one another, this is based on the standard error of estimate for the BFIHOST model (Boorman, *et al*, 1995). The distance threshold of 50km represents an empirical trade off between identifying a sufficiently large pool of candidate catchments, and

identifying catchments that are still close enough to the target catchment such that the two catchments experience similar climate patterns. Analogues resulting in duplicate pairs were rejected. For example, if catchment B was selected as the analogue for catchment A then catchment A was excluded from being selected as the analogue for catchment B.

The differences in area, catchment SAAR(61-90), AARO(61-90) and actual BFI between the target and the corresponding pool of potential analogue catchments were calculated for each target catchment. Where applicable, those potential analogues that were connected to the target catchment were also identified. The target catchments were then subdivided into two groups: those target catchments for which there were connected analogues (the connected group) and those for which all analogues were unconnected (the unconnected group). The best analogue catchment was selected for each target catchment within a group by selecting the analogue catchment with the smallest difference in catchment area. The connected group, selected in this manner, contained 42 target catchment pairs and the unconnected group contained 125 catchment pairs.

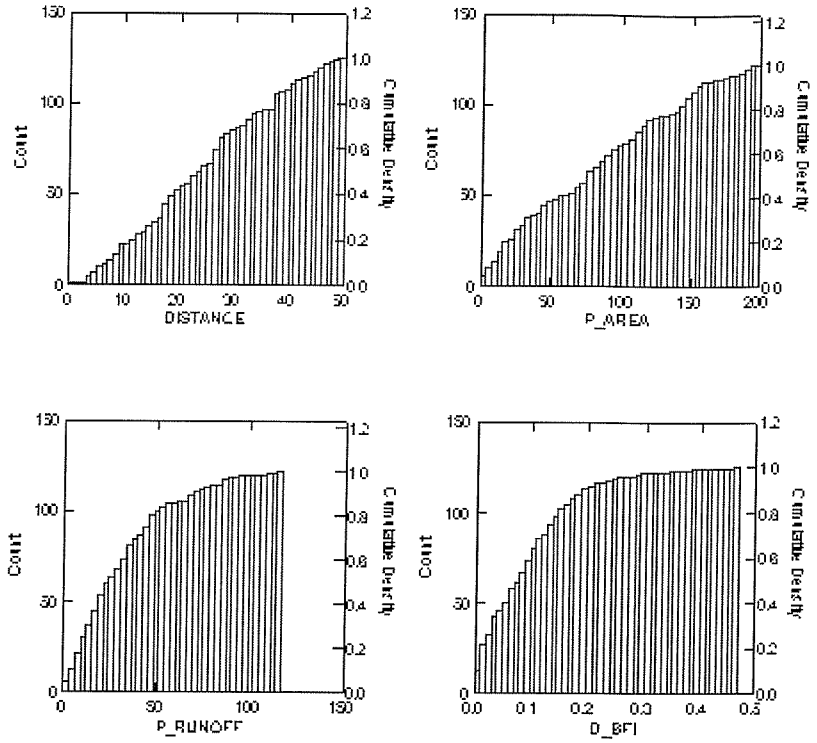
The characteristics of the catchment data set for MODA, the target set, are discussed in Chapter 7. For each catchment pair the similarity of analogue and target catchment areas and AARO(61-90) estimates was assessed by using distance measures representing the difference between the characteristic as a percentage of the smallest value of the characteristic for the pair. This statistic is respectively termed P_AREA and P_RUNOFF. Cumulative histogram plots for the geographic distance (DISTANCE) between pairs, P_AREA, P_RUNOFF and the difference in actual BFI are presented in Figure 9.1a for the unconnected data set and Figure 9.1b for the connected data set. For the unconnected data set these plots demonstrate that the distance and P_AREA statistics are fairly uniformly distributed between limits of [0,50] km for distance and [0,200] percent for P_AREA. The distributions for P_RUNOFF and the difference in actual BFI are skewed towards smaller values. It is worth noting that only 60% of catchment pairs have a difference in actual BFI of less than 0.1, which was the selection criterion for judging hydrological similarity on the basis of BFIHOST.

The patterns observed within the unconnected data set are generally also seen within the connected data set, although there are many fewer catchment pairs within the latter data set. The exceptions are that the majority of catchment pairs have a difference in BFI of less than 0.1 and that P_RUNOFF covers a much smaller range.

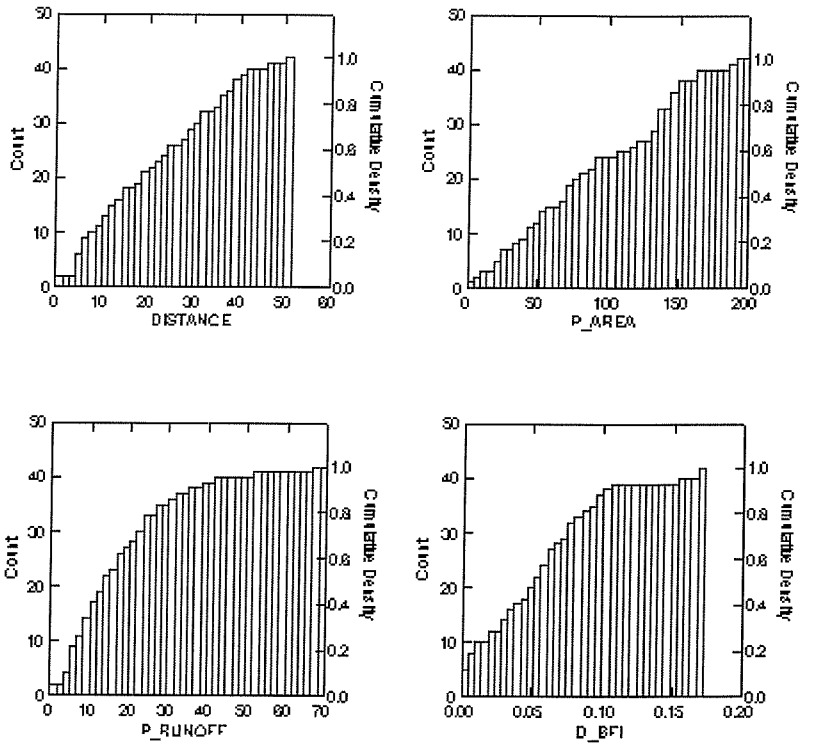
The three methods under evaluation were applied to each catchment pair over the period used for the calibration of MODA (Chapter 6). This calibration period was typically the 10 years leading up to and including 1997. This period was selected as the nearly all of the first choice analogues had flow data covering this period and, furthermore, the selection of the period meant that direct comparisons could be made with the results obtained with the regionalised form of MODB (Chapter 8). The results obtained for all catchments are summarised statistically in Table 9.1 for the unconnected group and within Table 9.2 for the connected group. For each group the distribution of the fit statistics obtained with each method over the catchments within the group are summarised as median and 68% upper and low limits. For the unconnected group, results obtained by re-scaling with catchment area and SAAR(61-90) are also presented for comparison with Method 2.

**Table 9.1 Analysis of transposition methods (Unconnected set)**

| Transposing Method | | Median | 68% l.l. | 68% u.l |
|---|---|---|---|---|
| Method 1 | A_BIAS | 7 | -30 | 59 |
| | A_EFF | 0.01 | -1.53 | 0.66 |
| | A_BEQ95 | 63 | -20 | 440 |
| Method 2 | A_BIAS | **3** | **-15** | **23** |
| | A_EFF | 0.62 | -0.09 | 0.81 |
| | A_BEQ95 | 5 | -42 | 156 |
| AREA | A_BIAS | 0 | -22 | 34 |
| &SAAR | A_EFF | 0.63 | 0.05 | 0.83 |
| | A_BEQ95 | 9 | -43 | 142 |
| Method 3 | A_BIAS | 1 | -57 | 261 |
| | A_EFF | **0.61** | **0.10** | **0.82** |
| | A_BEQ95 | **11** | **-35** | **122** |




**(a) Unconnected data set**




**(b) Connected data set**

**Figure 9.1 Distance histograms for catchment pairs.**

The method that gave the best fit is highlighted for each statistic in both tables. Considering the fit estimates for the unconnected data set, the smallest 68% c.i. for A_BIAS is obtained using Method 2 and the largest 68% c.i. is obtained using Method 3. Method 2, rescaling by estimated mean flow, offers a substantial improvement over solely rescaling by area (Method 1). Re-scaling by area and SAAR is an improvement over rescaling by area only, but it is not as efficient as re-scaling by estimated mean flow. Re-scaling by estimates of mean flow is an attempt to take into account the non-linearity between differences in rainfall and the resultant change in runoff. The results obtained using Method 3 are interesting, as re-scaling by estimated mean flow is implicit within this method. The much larger confidence interval for A_BIAS is probably a function of differences in the timing of flow between these unconnected pairs, particularly at higher flows percentiles.

Considering now the time series fit, as measured by the EFF statistic. Method 1 has a substantially larger 68% c.i. than the other methods. The median value of 0.01 is also very low, this indicates that, within 50% of catchments the mean of the observed flow data would represent a better fit than the use of the transposed data. Methods 2 and 3 and re-scaling by area and SAAR gave very similar results in terms of the upper 68% confidence limit and the median values, however Method 3 does have a marginally higher 68% lower limit that the other methods. The 68% lower limit for re-scaling by area and SAAR is higher than that for Method 2, although all of the methods have low 68% lower limits. A similar pattern is seen in the bias at Q95 flows (A_BEQ95), with Method 3 marginally outperforming the methods that take into account differences in rainfall regime as well as catchment area. Again re-scaling by catchment area alone results in a large confidence interval, and a marked tendency to over predict at the Q95 flow for the ungauged site.

These results indicate that the lowest errors in predicted mean flow are obtained when using Method 2 but, in terms of the time series fit there is a marginal advantage in using Method 3. However, this advantage does not offset the much larger errors at mean flow seen with the use of Method 3.

Considering the results for the connected data set, all three methods gave much better fits when applied to the connected data set than the unconnected set. Poorer fits were obtained

with Method 1 than with the other methods, however the use of this method still resulted in better fits than those obtained with any of the methods on the unconnected data set, although it should be remembered that the data set is much smaller. In the connected data set, the best fits are obtained when using Method 2 for all three measures of fit. The marginal improvements in the time series fit (as measured by A_EFF and A_BEQ95) observed in the un-connected set, when using Method 3 are not seen within this data set. This indicates that, when the target and analogue catchments are connected, the strong serial correlation between the flows measured at the two reaches ensures a good time series fit without the need for any need for the marginal improvements Method 3 gives in the unconnected data set. Indeed, Method 3 tends to yield poorer time series fits than Method 2, particularly for the bias error at Q95. This is probably a result of the uncertainty associated with the estimates of the flow duration curves from catchment characteristics, and hence in the estimation of correction factors. The factorial standard errors for the flow duration curve estimation model range from 1.08 at Q20 to 1.8 at Q99 (Holmes and Young, 2000b).

**Table 9.2 Analysis of transposition methods (Connected set)**

| Transposing Method | | Median | 68% l.l | 68% u.l. |
|---|---|---|---|---|
| Method 1 | BIAS | 8 | -12 | 28 |
| | Efficiency | 0.86 | 0.67 | 0.95 |
| | BEQ95 | 20 | -12 | 56 |
| Method 2 | BIAS | **3** | **-6** | **13** |
| | Efficiency | **0.89** | **0.72** | **0.96** |
| | BEQ95 | **13** | **-4** | **49** |
| Method 3 | BIAS | 85 | -24 | 496 |
| | Efficiency | 0.89 | 0.51 | 0.96 |
| | BEQ95 | 20 | -14 | 106 |

Over the two data sets, the best overall method for transposing analogue data is Method 2. Pearson correlation analysis was employed to look for correlation between the quality of fit (as measured by the BIAS, EFF and BEQ95 fit statistics), the characteristics of the target catchment (catchment area (AREA), estimated runoff (AAR) and estimated BFI, (BFIHOST)) and the distance measures for the distance of the analogue from the target catchment (P_AREA, DISTANCE, P_RUNOFF, D_BFIHOST).

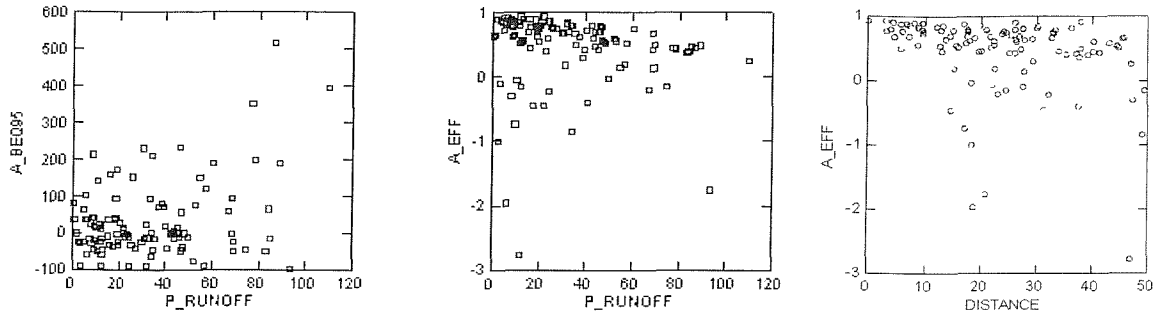



**Figure 9.2 Relationships between fit statistic and distance measures for the unconnected set.**

For the unconnected data set, the only significant correlation (at the 95% confidence level) were ones of –0.62 and 0.34 between EFF, BEQ95 and P_RUNOFF and one of 0.4 between EFF and DISTANCE. These relationships are graphed in Figure 9.2. These graphs demonstrates that EFF decreases as the percentage dissimilarity in runoff (P_RUNOFF) between the catchment pairs increases, although cases of very low efficiency where the mean of the observed would give a better fit (less than 0) occur across the entire P_RUNOFF range. The relationship between BEQ95 and P_RUNOFF is restricted to gross errors occurring only at high values of P_RUNOFF. The relationship between EFF and DISTANCE shows that efficiency decreases with increasing distance, and that efficiencies of less than zero are not observed for catchment pairs that are less than 13km apart. All of these correlations are likely to be related to the similarity of climatic regimes, although it should be noted that P_RUNOFF and DISTANCE are not significantly correlated.

For the connected data set the only significant correlation was one of –0.64 between P_RUNOFF and EFF. The pattern of this relationship, presented in Figure 9.3, was the same as that seen for the uncorrelated data set except that, with two exceptions, the efficiency values are all greater than 0.25.

This simple correlation analysis demonstrated the importance of the similarity of the climatic regimes experienced by the target and analogue catchment in determining the time series fit of the transposed data. Furthermore, this confirms that a good analogue catchment is one that is close, similar in terms of climatic regime and that is hydrogeologically similar, and thus has a similar runoff response to rainfall.

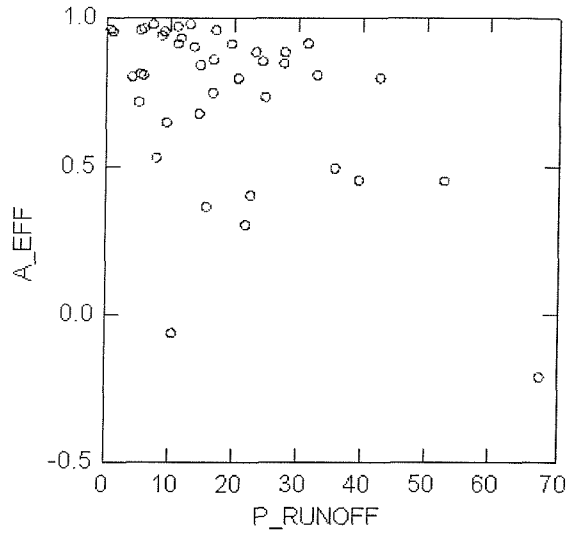



**Figure 9.3 Relationship between P_RUNOFF and the Efficiency statistic for the connected set.**

### 9.1.3 Summary

The evaluation studies have demonstrated that, of the methods tested, transposing gauged from data from an analogue catchment on the basis of a estimate of catchment mean flow, derived from the Low Flows 2000 regional model for predicting mean flow is the most efficient method for transposing data. This method, Method 2, gave the best results within the connected catchment pairs and provided the best trade off between simulating the time series structure of the flow data at the target site and maintaining an acceptable water balance within the unconnected catchment pairs. Correlation analysis has demonstrated the importance of the similarity of the climatic regimes experienced by the target and analogue catchment in determining the time series fit of the transposed data. The evaluation has confirmed that a good analogue catchment is one that is close and similar in terms of climatic regime and hydrogeology to the target catchment. Furthermore, the transposing of data from connected pairs results in a good simulation of the times series structure of the flow data at the target site for the majority of cases. This is in contrast with the case where the target and analogue catchments are unconnected, in this instance the quality of the time series simulations obtained varies widely in terms of the fit statistics used.

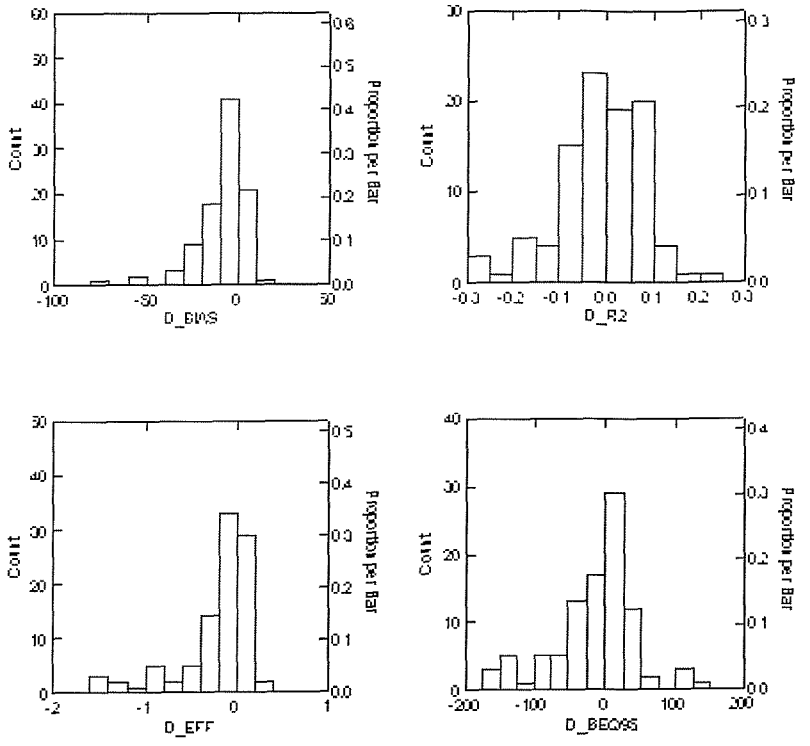
## 9.2 COMPARISONS WITH REGIONALISED MODB MODEL FITS

Section 9.1 presented an evaluation of methods for transposing gauged flow records from suitable analogue sites. An important issue to be considered is the conditions under which a transposed gauged flow record will result in a better simulation at a target site than the application of a regionalised rainfall runoff model. To evaluate this, a comparison was made between the quality of the fit statistics (BIAS, EFF and BEQ95) obtained using MODB with regionalised model parameters with that obtained by transposing gauged flow data using Method 2. MODB was used as the analysis of model fit described in Chapter 8 indicates that the model fits obtained with the regionalised form of MODB were better than those obtained with the regionalised form of MODA.

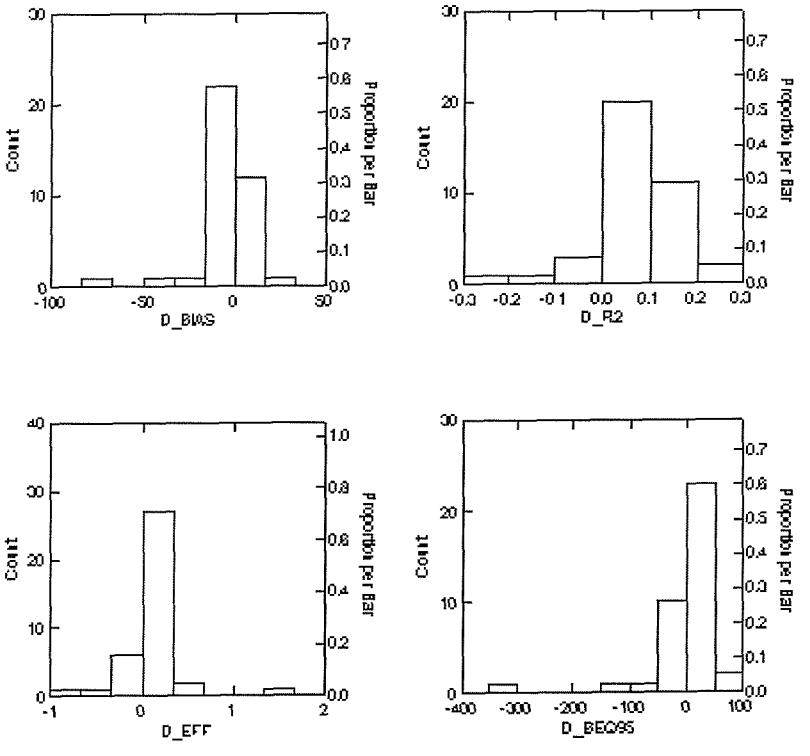
The differences between the fit statistics obtained using the analogue Method 2 and those obtained using the regionalised model were used to compare the two approaches. Differences were calculated such that a negative result indicates that the regionalised rainfall runoff model yields a better result that the use of transposed data from the analogue gauge. This analysis considered the calibration period originally used for the rainfall runoff modelling work. Histograms of the differences between fit statistics are presented in Figure 9.4a for the unconnected data set and Figure 9.4b for the connected set. For the unconnected set there were 110 catchments for which MODB estimates were available, and for the connected data set there were 30 catchments. In these figures the statistics are prefixed by "D_" to denote that the differences in the statistics are being graphed. Within the unconnected data set there were a number of catchments that gave significantly poorer results than the regionalised rainfall runoff model for one or more statistics. These catchments were screened out for graphing purposes. The screening criteria used were:

- differences in bias (D_BIAS)> -100;
- differences in efficiency (D_EFF)> -2;
- differences in Bias at Q95 (D_BEQ95)> -500.

The catchments pairs removed using these screening criteria are summarised in Table 9.3, together with the criterion that led to them being excluded. The actual BFI and catchment SAAR(61-90) values for the target catchment are also presented within the table.




**(a) Unconnected data set**




**(b) Connected data set**

**Figure 9.4 Difference histograms for fit statistics.**

**Table 9.3 Catchments pairs excluded from the histograms for the unconnected set**

| Exclusion criteria | Catchments | | Target catchment Characteristics | | | Regionalised MODB fit statistics | | | Analogue method Fit statistics | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | target | source | Area $km^2$ | BFI | SAAR (mm) | BIAS | EFF | BEQ95 | BIAS | EFF | BEQ95 |
| D_BEQ95<-500 | 6008 | 7001 | 106 | 0.30 | 1291 | 3 | 0.67 | 371 | 20 | 0.53 | 1166 |
| | 36009 | 37011 | 26 | 0.32 | 598 | 3 | 0.61 | 1309 | 82 | **0.79** | 2094 |
| | 21023 | 21012 | 113 | 0.35 | 671 | 5 | 0.61 | 290 | 18 | 0.16 | 1034 |
| | 39054 | 41029 | 34 | 0.24 | 816 | 25 | 0.72 | 92 | 298 | -5.73 | 654 |
| | 46818 | 47009 | 3 | 0.38 | 1250 | 5 | 0.71 | 566 | -7 | 0.59 | 1080 |
| | 27051 | 27043 | 8 | 0.45 | 855 | -1 | 0.73 | -9 | 3 | 0.46 | 515 |
| D_EFF<-2 | 23004 | 23008 | 750 | 0.33 | 1147 | 2 | 0.70 | -10 | 177 | -5.56 | 208 |
| | 39016 | 39065 | 1033 | 0.87 | 759 | 16 | 0.61 | -33 | -99 | -1.76 | -98 |
| | 52010 | 52011 | 140 | 0.47 | 867 | 0 | 0.70 | -7 | 221 | -10.58 | 58 |
| | 27055 | 27057 | 131 | 0.69 | 882 | -8 | 0.55 | -57 | **3** | -1.96 | -61 |
| | 51001 | 52007 | 74 | 0.67 | 911 | 1 | 0.90 | -24 | 35 | -2.78 | **11** |
| | 40007 | 41029 | 252 | 0.48 | 830 | 18 | 0.68 | -56 | 218 | -13.65 | **-1** |
| D_BIAS_100 | 55035 | 55033 | 1 | 0.29 | 2467 | -2 | 0.79 | -38 | 145 | **0.93** | **13** |

Those catchments with large BEQ95 values and/or poor general time series fits (as judged by the values of EFF) all tend to be low rainfall catchments. Significant soil moisture deficit might be expected to occur during the summer months in these catchments. The catchments with very poor BEQ95 values also tend to be relatively small catchments. The catchment excluded on the basis of BIAS error is an extremely small, wet catchment. Catchment pairs that were excluded on the basis of one fit statistic, but where transposition gave a better result than the rainfall runoff model for one or more of the other fit statistics are highlighted within the table. In the majority of catchments the fit obtained by transposing the data from an analogue gauge gave a poorer fit for the other statistics, as well as the excluding statistic. Within the one catchment that failed on the BIAS estimate, the analogue method gave very good fits for both efficiency and BEQ95. These were better than the values obtained using the regionalised rainfall runoff model.

Considering the histograms in Figure 9.4b for the connected data set. In the majority of cases the quality of the time series fit (EFF) and fit at low flows (BEQ95) obtained by transposing gauged flow data was generally better than that obtained using the regionalised rainfall runoff model. This is not the case for overall BIAS (error in mean flow), in this case the use of the regionalised rainfall runoff model generally gave better results. Of the 30 catchment considered, the use of the analogue data gave higher values of efficiency in 27 catchments with an average improvement of 0.19 over the values obtained with the

regionalised rainfall runoff model. This is reduced to 17 catchment pairs for BEQ95 with an average reduction in bias error at low flows of 23%. For overall BIAS only 11 of the 30-catchment pairs gave a better result than that obtained using the regionalised rainfall runoff model, with an average reduction in BIAS of 4%. In the 19 catchments, where the use of the regionalised rainfall runoff model gave a smaller overall BIAS the average improvement in BIAS was 10%.

Considering the unconnected data set, Figure 9.4b, the use of the regionalised model generally gave better results for all fit statistics. Of the 110 catchment pairs considered, the use of analogue data gave better results in only 23 catchments with respect to a reduction in BIAS, 33 for an improved efficiency and 50 for a reduction in bias at Q95. The average improvements were 4%, 0.18 and 18% respectively. These figures are much smaller than the average improvements obtained when the use of the regionalised rainfall runoff model gave better results. These improvements were 14% for BIAS, 0.25 for EFF and 165% for the BEQ95. The last figure is biased by a small number of catchments in which very poor fits were obtained when using data from the analogue catchment.

This analysis has demonstrated that, in that majority of cases, if the analogue catchment is connected to the ungauged catchment, and both are natural the transposing of gauged flow data generally yields a better time series fit that that obtained through using a regionalised rainfall runoff model. However, the use of a regionalised rainfall runoff model will generally give a marginally better water balance. These conclusions are subject to the condition that a suitable analogue catchment, meeting the selection criteria used, can be identified. In the case where the analogue and ungauged catchments are not connected, the use of a regionalised rainfall runoff model will generally give better results than the use of data from the analogue site. Furthermore, the use of analogue data can result in very large errors, both from the perspective of replicating the time series nature of the data and the closing of an adequate water balance. The good replication of the time series nature of the flow data for the connected data set is a consequence of the strong serial correlation between the flows observed at the target site and the analogue site.

## 9.3 THE PREDICTION OF THE APPROPRIATE ESTIMATION METHOD

The analysis in the previous section considered each of the flow statistics independently. From a practical viewpoint, it is important to consider whether it is feasible to predict under what conditions the use of a regionalised rainfall runoff model will give a better result than that which would be obtained by transposing measured data from the most appropriate analogue catchment. In order to provide guidance on this issue it is necessary to consider the fit statistics as a group of statistics that describe the quality of a simulated time series, rather than as individual measures. To address this prediction issue, discriminant analysis has been used to evaluate whether the characteristics of the target catchment and the distance of the analogue from the characteristics of the target catchment can be used to determine which method is more appropriate for a given set of conditions. The basic data used for this investigation were the sets of target-analogue catchment pairs used within the comparison work, described in the previous section.

Initially, Pearson correlation analysis was used to look for relationships between the differences in the goodness of fit statistics for the two data sets and the characteristics of the target catchment and the distance measures between the catchment pairs. The variables entered for the target catchment were AREA, AARO(61-90), BFIHOST and the distance measures were P_AREA, P_RUNOFF, P_BFIHOST and DISTANCE. The goodness of fit statistics were BIAS, EFF and BEQ95.

There were no significant correlations identified for the connected data set. For the unconnected data set, significant negative correlations of -0.3, -0.6 and –0.3 were identified between P_RUNOFF and the differences in BIAS, EFF and BEQ95. The correlations, confirmed by scatter plots, indicate that the use of analogue data can give better results that the use of the regionalised rainfall runoff model if the difference in estimated runoff between the two catchments is small. There was a significant positive correlation of 0.3 between the estimated runoff for the target catchment and the differences in BIAS. This correlation indicates that the use of analogue data can give a better result than the use of a regionalised rainfall runoff model in high rainfall, and hence high runoff catchments. In these catchments significant soil moisture deficits are not generated and evaporation will tend to take place at the potential rate. Under these conditions, transposing data from analogue catchments with

similar climatic regimes is more likely to give acceptable water balances as the catchment losses will not be as sensitive to, potentially, subtle differences between the rainfall-runoff regimes of the two catchments. There was also a negative correlation of -0.3 between DISTANCE and D_EFF indicating that good time series fits can be obtained using analogue data when the distance between the catchments is small, and hence the climatic regime (particularly with respect to rainfall patterns) is likely to be similar.

The frequency of obtaining a better fit for more than one of the statistics, for a given method, was assessed for both data sets. The number of catchments in which a method for estimating flow at the target site gave significantly better results for BIAS and EFF, and BIAS, EFF and BEQ95 is presented in Table 9.4 for the analogue Method 2 and the regionalised form of MODB. The thresholds for determining whether one method gave a better fit than the other for each statistic were:

- the reduction in the magnitude of BIAS had to be greater than 5%;
- an increase in efficiency had to be demonstrated;
- the reduction in the magnitude of the bias at Q95 had to be greater than 10%.

These thresholds are somewhat arbitrary in nature, but the thresholds do reflect the hydrometric uncertainty with the measurement of the statistics and the degree of artificial influence allowed for in the original catchment selection criteria (Chapter 5). The results from this analysis are summarised in Table 9.4. The last line in the table summarises the number of catchments where neither method gave a better result for all the statistics concerned.

For the unconnected data set, the numbers within this table demonstrates that the rainfall runoff method give better results for BIAS and EFF in 66% of the catchments and gave better results for all three statistics within 48% of catchments. The analogue method does not perform well across two or more statistics within the unconnected data set, with the number of catchments being lower than the number of catchments where mixed results were obtained. This situation is reversed across the connected data set, in this set the analogue method is significantly better than the regionalised model. The number of catchments where the regionalised model out performed the analogue method is very small.

**Table 9.4 Catchments classified according to two or more fit statistics**

| | Unconnected set | | connected set | |
|---|---|---|---|---|
| | BIAS and EFF | BIAS, EFF and BEQ95 | BIAS and EFF | BIAS, EFF and BEQ95 |
| **Analogue method 2** | 15*(14%)* | 9 *(8%)* | 18 *(60%)* | 15 *(50%)* |
| **MODB (regionalised)** | 73 *(66%)* | 53 *(48%)* | 3 *(10%)* | 2 *(7%)* |
| **Mixed results** | 22 *(20%)* | 48 *(44%)* | 9 *(30%)* | 13 *(43%)* |

Discriminant analysis was used on the unconnected data set to identify whether the catchment pairs can be classified on the basis of the fit statistics. This analysis was not applied to the connected data set as the regionalised model outperformed the analogue Method 2 in only a few catchments. For the unconnected data set the individual catchment pairs were coded according to whether the best fits for BIAS and EFF were obtained with; the analogue methods (code=1), the use of a regionalised rainfall runoff model (code=2) or whether the results were mixed (code=0).

The first step in the analysis was to use backwards stepping discriminant analysis to identify which of the variables entered were useful in explaining the variation between groups. The test for usefulness was the F statistic, this statistic is based on the ratio of the between group variation in the mean value of the variable to the mean of the within group variation between members and the group mean. The larger this statistic is for a variable, the more useful the variable is in determining the classification of group members. The threshold for removal of a variable was an F statistic of 3.9. This choice, which equates to a mean within group variance equal to 25% of the between group variance is a recommended threshold (Systat, 1998). The analysis was not sensitive to small variations in the value of the F to remove statistic. This process confirmed the results from the Pearson correlation analysis that the predicted BFI value for the target catchment (T_BFIHOST), P_RUNOFF and DISTANCE were all useful in differentiating between groups.

A multi-layer scatter plot matrix showing the relationships between the classes and the differentiating variabless is presented in Figure 9.5. The ellipses are the 68% bivariate normal confidence intervals for the differentiating variables for each class. This plot demonstrates that there is a high degree of overlap between groups for all differentiating variable pairs. The plots do demonstrate that the analogue method is most effective in flashy catchments (low T_BFIHOST), where the difference in runoff between the analogue and target catchments is

small (P_RUNOFF) and the two catchments are in close proximity to one another (small DISTANCE).

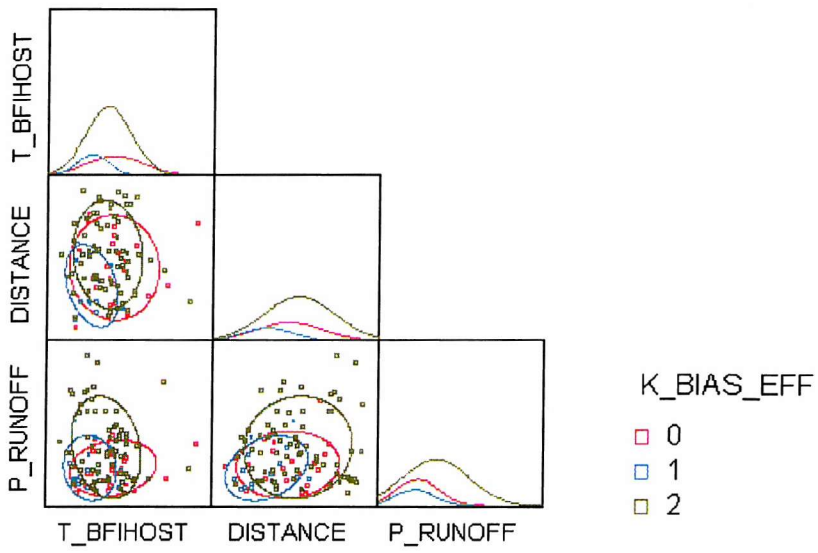



**Figure 9.5 The relationships between classes of fit and the classifying variables.**

The within group mean values of variables, and the mean F statistics for each group mean compared to it's neighbouring groups are presented in Table 9.5 and Table 9.6. These demonstrate that groups 1 and 2 are equally distinct from the mixed group, but these groups are more distinct from one another.

**Table 9.5 Group mean values**

| | Class Mean | | |
|---|---|---|---|
| **Variable** | **Class 0** | **Class 1** | **Class 2** |
| T_BFIHOST | 0.52 | 0.41 | 0.49 |
| DISTANCE (km) | 21 | 15 | 26 |
| P_RUNOFF % | 21 | 21 | 38 |

**Table 9.6 F statistics for the groups**

| **Class** | **0** | **1** | **2** |
|---|---|---|---|
| **0** | 0.00 | | |
| **1** | 3.92 | 0.00 | |
| **2** | 3.92 | 8.24 | 0.00 |

The classification functions obtained from the discriminant analysis are presented in Table 9.7. The classification of target catchments according to the classification functions (the class assigned is the one for which the function has the largest value) is summarised in Table 9.8. The jack-knifed classification of catchments is presented in Table 9.9. A jack-knifed classification is one where the classification functions are re-evaluated for each catchment assignment with the catchment to be assigned omitted from the data set. These tables present the actual class for catchment in rows and the class assigned on the basis of the classification function in the columns. For each class, there is also a column summarising the percentage of catchments that are correctly assigned to classes.

The jack knifed sample assignments do not vary that much from those obtained using the full classification functions. This indicates that the analysis is not over specified, and that the classification functions are stable. Looking at the full classification, 61% of catchments are classified correctly. The percentage is highest for Class 1 and lowest for Class 0, the mixed result class. Where catchments are incorrectly classified, the allocation of these catchments to the remaining classes is fairly random because of the large overlap between all classes.

The classification of catchments to classes 0 and 1 is better than that which would be obtained by chance (20% and 14% respectively (Table 9.4)), however the classification of catchments to Class 2 is lower than that which would obtained by chance (66%). Furthermore, 20% of Class 2 catchments are incorrectly assigned to Class 1.

**Table 9.7 Group classification functions**

| | Function coefficients for Classes | | |
|---|---|---|---|
| **Variables** | **Class 0** | **Class 1** | **Class 2** |
| CONSTANT | -14.811 | -9.588 | -15.530 |
| T_BFIHOST | 43.428 | 34.307 | 41.843 |
| DISTANCE | 0.165 | 0.117 | 0.195 |
| P_RUNOFF | 0.058 | 0.053 | 0.085 |

**Table 9.8 Classification Matrix**

| Class "n" | Class assignments of catchments from Class "n" | | | % assigned |
|---|---|---|---|---|
| | Class 0 | Class 1 | Class 2 | Correctly |
| **Class 0** | 12 | 6 | 5 | 52 |
| **Class 1** | 2 | 11 | 2 | 73 |
| **Class 2** | 13 | 15 | 45 | 62 |
| **Total** | 27 | 32 | 52 | 61 |

**Table 9.9 Jack-knifed classification matrix**

| Class "n" | Class assignments of catchments from Class "n" | | | % assigned |
|---|---|---|---|---|
| | Class 0 | Class 1 | Class 2 | Correctly |
| **Class 0** | 10 | 6 | 7 | 43 |
| **Class 1** | 2 | 11 | 2 | 73 |
| **Class 2** | 17 | 15 | 41 | 56 |
| **Total** | 29 | 32 | 50 | 56 |

To represent the overlap between classes, the discriminant analysis was repeated using canonical classification functions. These are factor functions that are orthogonal to one another and which sequentially maximise the possible F ratio for the variation within, and between groups. There are m-1 functions identifiable from the data, where m is the number of classifying variables - in this case three. The plots of classified catchments against the two functions identified from the data are presented in Figure 9.6. Bi-variate normal 68% confidence interval ellipses are also plotted within this figure for each class. This plot clearly demonstrates the overlap between classes.

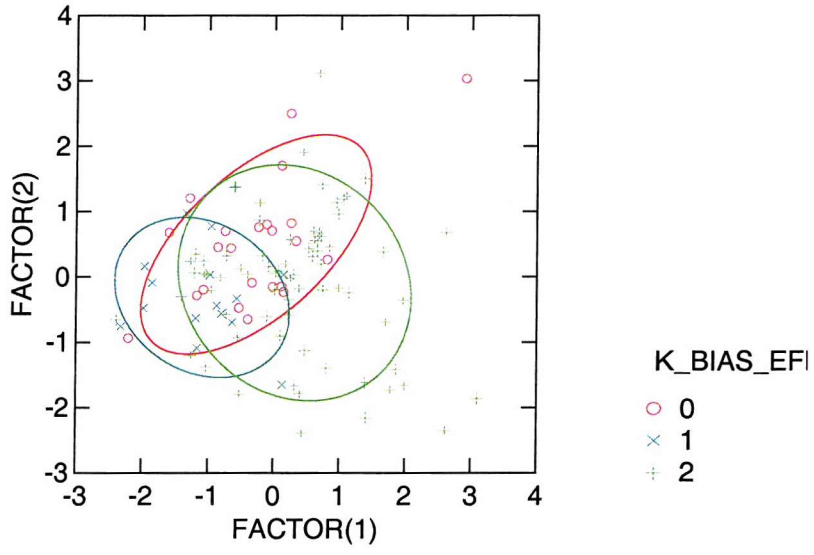



**Figure 9.6 Canonical scores plots for catchment assignments.**

## 9.4 SUMMARY

The evaluation of transposition methods has demonstrated that the most efficient method for transposing gauged data from an analogue catchment is one based on an estimate of catchment mean flow derived from the Low Flows 2000 regional model for predicting mean flow. This method gave the best results within connected catchment pairs, and provided the best trade off between simulating the time series structure of the flow data at the target site and maintaining an acceptable water balance within unconnected catchment pairs. The evaluation analysis also demonstrated the importance of the similarity of the climatic regimes experienced by the target and analogue catchments in determining the time series fit of the transposed data. The evaluation studies have confirmed that a good analogue catchment is one that is close, similar in terms of hydrogeology and climatic regime, and thus has a similar response to rainfall inputs to that of the target catchment. Furthermore, the analysis has demonstrated that transposing of data from a connected analogue resulted in a very good simulation of the times series structure of the flow data at the target site for the majority of cases that met the catchment selection criteria. This contrasts with the case where the target and analogue catchments are unconnected, in this instance the quality of the simulations, in terms of the fit statistics used, obtained by transposing the data from the analogue catchment was often poor.

The quality of fit obtained by transposing data from a "best" analogue catchment was compared with the fit obtained by applying the regionalised form of the MODB rainfall runoff model. The analysis demonstrated that, in the majority of cases, if the analogue catchment is connected to the ungauged catchment, and that both are natural, the transposing of gauged flow data generally yields a better time series fit that that obtained through using a regionalised rainfall runoff model. However, the use of the regionalised rainfall runoff model generally gave a marginally better water balance. The good replication of the time series nature of the flow data for the connected data set is a consequence of the strong serial correlation between the flows observed at the target site and the analogue site.

In cases where the analogue and ungauged catchments are not connected, the use of a regionalised rainfall runoff model will generally give better results than the use of data from the analogue site. Furthermore, the use of flow data from an unconnected analogue can result in very large errors, both from the perspective of replicating the time series nature of the data and the closing of an adequate water balance.

Discriminant analysis was used to investigate whether it is possible to predict when the use of a regionalised rainfall runoff model will yield a better overall stream flow simulation that that obtained by transposing data from an analogue catchment. The BIAS and EFF statistics were used to define what constituted a better simulation. A better simulation for a method was defined as one that gave an improved model fit for one statistic, compared with the other method, and a model fit that was at least as good as the fit obtained using the other method for the other statistic. The discriminant analysis was only applied to the unconnected set due to small sample size available for the connected data set. Within the connected data set the use of analogue data generally gave a better time series fit than the use of a rainfall runoff model whereas the regionalised rainfall runoff model was often marginally more effective in closing a water balance.

The application of discriminant analysis to the unconnected set considered the characteristics of the target catchment and the runoff, area and geographic distance measures for discriminating between simulation classes. The results from the analysis demonstrated that, on the basis of these characteristics and distance measures, it was not possible to develop a robust method for predicting which method is most appropriate. This is primarily due to the large overlap of classes. However, in the unconnected site there is a much greater chance that the use of the rainfall runoff model will give a better simulation of stream flow than the use of the analogue method.

The catchments used in the original rainfall runoff modelling calibration work (Chapter 5) have been used as the basis of the comparison analyses presented within this Chapter. These catchments are natural, of good hydrometric quality, and the climatic data and the definition of catchment boundary used within the modelling work have been demonstrated to be acceptably free from error. The conclusions from the comparison studies must be viewed in the context of a comparison where the climatic data required for applying the regionalised

rainfall runoff models can be assumed to be without error. The flow data for the analogue catchment also has to be assumed to be both natural and without error. The catchment area defined by the catchment boundary is used in both methods, and thus the overall results must be viewed in the context of being able to define an accurate hydrological catchment boundary for both methods. In the context of the analogue method, this consideration also applies to the analogue catchment.

# 10 Conclusions

## 10.1 ACHIEVEMENTS AND CONCLUSIONS

The overall research objective of this thesis has been to develop a rainfall runoff model for predicting the natural variation of daily stream flows within a catchment without recourse to the calibration of model parameters against observed stream flow data. As discussed within Chapter 1, there is an implicit requirement within this objective that the model parameters can be estimated from the physical characteristics of the catchment.

The thesis is the largest study of this type within the UK, and is also larger than any international studies reported within the scientific literature. The study has focused on data errors, model structure and calibration and subsequent regionalisation strategies. Through this approach, it is the first to have demonstrated that, in the UK context, a conceptual model structure for predicting daily stream flows can be defined such that conceptually justifiable relationships can be derived between model parameters and the physical characteristics of the catchment being modelled. Furthermore, it has been demonstrated that the errors in the simulated stream flows are sufficiently small for the techniques to be a useful aid in the management of water resources within the UK.

The fundamental approach to the research was to calibrate suitable models on stream flow within a large, representative sample of catchments and to subsequently relate the model parameters to the climatic and physiographic characteristics of the catchments. As discussed in Chapter 6, the parameters of a calibrated model are a function of the:

- input data (including observed stream flow) and errors within the input data;
- the model structure;
- structure of the catchment being modelled;
- the calibration scheme (including the selection of objective functions).

The chapters of this thesis have been structured as a logical progression from the evaluation and development of techniques for estimating requisite input data, through to the final evaluation of the success in meeting the overall research objective. The achievements of the research within each chapter are discussed fully within the chapter concerned. This chapter brings together the key achievements and the increase in the hydrological knowledge base from each research area.

The history of international research within this area was reviewed in Chapter 1. The chapter highlighted that previous studies had been restricted to relatively small sample sets of catchments. Commonly previous studies have been directed at simulating single events or monthly stream flow and/or annual runoff. The reason for this was either a function of requirement, or a practical consideration based on data availability and/or a desire to not consider the detailed hydrological processes that have to be considered if the objective is to simulate daily stream flows. The study of Sefton and Boorman (1997) which reported on the regionalisation of daily resolution models within the UK was inconclusive, a consequence of the small sample size considered (less than 40 catchments) and the inconclusive nature of the derived relationships between model parameters and catchment characteristics.

The literature on the impacts of errors in climatic input data on model behaviour was reviewed within Chapter 2. The chapter presented an evaluation of derivatives of the British Standard methods for the areal estimation of daily rainfall time series. The analysis highlighted the sparseness of rain gauge data within the north and upland areas of the UK. This was found to impact upon the accuracy of the daily variation in rainfall but not to unduly impact upon the accuracy of the estimation of annual rainfall totals, if information from long term average annual rainfall maps was included within the estimation procedure. There was some evidence that annual rainfall totals were more poorly estimated during "dry" years. It is in these years, particularly in the south and east of the UK, that the correct modelling of resultant soil moisture deficit is critical for accurately modelling the stream flow response of a catchment to precipitation and closing an adequate water balance.

Chapter 2 also presented the developed of a scheme for dis-aggregating MORECS Penman Montieth PE estimates from a 40 to a 1 km resolution grid taking into account the impact of increased spatial heterogeneity of topographic controls on PE. This addresses the major criticism of the use of MORECS in rainfall runoff modelling.

If a daily rainfall runoff model is to be successfully regionalised, from a water resources perspective, it is essential that the model, when calibrated, can accurately simulate the important aspects of the observed stream flow within a catchment. Furthermore, it is essential that the model is stable when evaluated against an independent period of stream flow data. From a water resources perspective, this implies that the model can close a water balance, that recession periods are modelled effectively and that the overall time series variation in stream flow is adequately modelled. It is therefore pre-requisite that a model can accurately simulate the behaviour of soil moisture in controlling evaporation losses, the generation of effective rainfall, drainage to groundwater if appropriate, and the correct hill slope routing of direct runoff. The model should also be physically realistic, such that the parameters are physically relevant. This latter consideration requires a compromise between the physical realism of the conceptual representation of process within the model and over-specification, which commonly results in structural errors.

The evaluation of a range of rainfall runoff model philosophies within five catchments in East Anglia was presented within Chapter 3. These catchments are very low rainfall catchments, thus necessitating the accurate simulation of soil moisture behaviour. A novel multi-function approach to evaluating combined model performance over both a calibration and evaluation period of record was also prototyped within this analysis. Based on the results of the model evaluation studies, two model structures, MODA and MODB were developed for application within a regionalisation scheme. The development of these models was presented within Chapter 4. In developing these models, attention was paid to the conceptual representation of the hydrological processes within the models. Both structures utilised the concept of a probability-distributed representation of soil storage capacities, developed and used by Moore (1985) amongst other researchers. Both models used a second order linear routing reservoir scheme for simulating quick and slow flow routing of effective rainfall. MODA, the more complex model, included an interception store and a soil moisture related drainage term, and had five free parameters for calibration.

MODB had three parameters and employed a fixed partitioning of effective rainfall between quick and slow flow routing paths, based on the soil classes within each catchment.

The development of a large catchment data set, with good quality gauged stream flow records and a comprehensive set of catchment characteristics and descriptors was presented within Chapter 5. This data set is the largest reported data set used within a rainfall runoff regionalisation study. Characteristic and descriptors for the catchments included: topographic descriptors of the catchment and stream network within the catchment, derived using a DTM; a range of climatic regime characteristics; characteristics describing the catchment soils (HOST, Boorman *et al*, 1995) and characteristics describing catchment vegetation classes (Fuller *et al*, 1994). It was demonstrated that the variation of the characteristics within the data set were representative of the variation of these characteristics across the UK.

The classical approach to calibrating a rainfall runoff model employs the use of an automatic search algorithm to search a feasible parameter space coupled with manual intervention. Manual intervention enables the hydrologist to make non-linear fuzzy decisions about the trade off between different aspects of model fit, based on experience. A review of automatic calibration schemes, problems associated with parameter identifiability and the impacts of uncertain input data were discussed in Chapter 6.

The development of a novel, global, search scheme was presented in Chapter 6. The objective was to develop a scheme that could automatically implement the sort of decisions that an experienced hydrologist would make as a set of fuzzy rules. The scheme developed is a constrained random walk scheme that utilises a number of objective functions, allows for trade off between different aspects of model fit and recognises that the input stream flow data have an associated uncertainty. Within a catchment, the scheme is used to identify a large number of parameter sets, or vectors, that can be considered equally likely given these constraints. The final selection of a model parameter vector, to represent the catchment with the regionalisation analysis, is made on the basis of the combined performance of the calibrated model for each vector over both the calibration period and the stability of the model when applied to an independent evaluation period. A combined

Euclidean objective function was developed for this selection. The components of the function assess the model fit from the perspective of closing a water balance, the general time series structure of the data and the fit over recession periods, with a particular emphasis on low flows. The selection of a best compromise parameter vector reflects that a good, stable model fit is more likely to be a function of the catchment type than one that just gives a good fit over a calibration period, and hence may be incorrectly specified.

This scheme also enabled the behaviour of the model to be evaluated within specific catchments. The results for four such catchments are also discussed in Chapter 6. This analysis identified that even a simple model such as MODA had potential structural problems giving rise to parameter covariance issues. This led to the development of MODB. These type of behaviour analyses identified how the model parameters compensated for errors in input data and also produced evidence that the importance of model components, and hence the identifiability of parameters, could potentially vary between catchment types. The choice of parameters in low rainfall catchments was found to be very sensitive to the requirement that the model should be capable of closing a water balance. This is an essential requirement from a water resources perspective, but is not one that is widely recognised within the literature of model calibration. The inability to close an adequate catchment water balance led to a number of potential catchments being rejected at this stage. A review of these catchments identified potential errors within rainfall data and the presence of groundwater, which makes the assumption of a catchment water balance questionable- as being the primary reasons for the exclusion of catchments.

An evaluation of the calibrated model performance (Chapter 7) identified that both models simulated stream flow extremely well in the majority of catchments. Over the calibration period, MODA was better at predicting mean flow and MODB was better at predicting low flows. Over the evaluation period, the model fits for MODB were generally more stable than those for obtained for MODA, also indicating that MODA was over specified.

Within Chapter 7 the development of conceptually supportable relationships between model parameters and catchment characteristics/descriptors is presented. These were developed using a multivariate regression approach for both MODA and MODB. Good or acceptable relationships were identified for all model parameters with the exception of one

of the soil store parameters for MODA and the sole soil storage parameter for MODB. The uncertainties in the relationships between model parameters and catchment characteristics are a function of the physical reality of the structure of the rainfall runoff model, errors in input data, calibration errors, error in catchment characteristic data and errors, or assumptions, in structure of the relationships themselves. The two parameters in question were identified in Chapter 6 as being sensitive to errors in climatic data (particularly precipitation), an effective estimate of contributing area and the assumption that any regional ground water flow does not bypass the gauging station. The poor relationships were attributed to the requirement to close a water balance during calibration. Extensive split sample tests were used to demonstrate that the regression models were not over specified, and hence were independent of the catchment sample used to develop the models.

The performance of the regionalised models was assessed in Chapters 8 and 9. In Chapter 8 a comparison was made between the regionalised and calibrated model forms across the catchment data set used for the study. Comparing the fit of the regionalised models with the calibrated models identified that over the calibration period the calibrated parameters resulted in much lower water balance errors - as would be expected as closing a water balance was a key part of the calibration scheme. Over the evaluation period, the performance of the regionalised models in closing a catchment water balance was very comparable to the calibrated models. A number of catchments in which the use of the regionalised model parameters gave very poor time series fits were identified. The calibrated models also performed less effectively in these catchments; a discussion of these catchments is presented in Chapter 8. However, over the majority of catchments, the time series fits obtained with the regionalised models were very comparable to those obtained with the calibrated models. This is a very significant result.

A classification of the ability of the regionalised models to both close an adequate water balance and maintain a reasonable hydrograph demonstrated that, over the evaluation period considered, MODB gave acceptable simulations in more of the catchments within the data set than MODA. For MODB simulations that were good or very good were obtained within 67% of catchments compared with 58% of catchments for MODA. MODB gave acceptable (or better) simulations in 85% of catchments compared within 78% of

catchments for MODA. The analysis demonstrated that the regionalised version of MODB was more effective in closing a water balance than the regionalised form of the more conceptually correct MODA.

Spatial analysis identified that catchments in which an acceptable water balance could not be closed tended to be located in the south of the country, and were dominated by groundwater catchments in the case of MODB. Catchments where the time series model fit was judged to be unacceptable tended to be dry catchments on the east of the country. Inspection identified that the rain gauge network was sparse over these catchments, potentially resulting in poor estimates of precipitation inputs to the models. The time series criterion for judging the acceptability of the simulation was harsh. The simulations for many catchments that failed on this criterion were able to close a water balance and simulated periods of recession well. The number of catchments that had completely unacceptable time series fits, from a water resources perspective, was considerably lower for both models than the classification suggested.

The regionalised model forms for MODA and MODB were applied to the two of the Anglian catchments from the model evaluation work presented within Chapter 3. These catchments were not used within the regionalisation research. The results obtained were compared within the results obtained for the four models calibrated, and evaluated, within these catchments. In one of the catchments the regionalised models gave a better simulation of stream flow over the evaluation period than that obtained using three of the four bespoke model applications. This again is a significant result supporting the overall conclusion.

The generally good performance of the regionalised models in predicting stream flow is a very significant result, and is not one that has been replicated in the literature on the regionalisation of rainfall runoff models. The study has demonstrated that the parameters were more identifiable and the model simulations were more stable for the three-parameter model (MODB) than the five-parameter model (MODA) in a calibrated application. This adds weight to the argument that, from a systems engineering point of view, only a very limited number of parameters may be identified from stream flow data (e.g. Beven, 1993; Jakeman and Hornberger, 1993). The regionalised form of MODB was demonstrably superior to the regionalised form of MODA in closing a water balance, despite the relative

conceptual simplicity of the model.

A comparison of the performance of the regionalised form of MODB and the current operational practice of transposing gauged from data from an analogue catchment was presented in Chapter 9. This included a rigorous evaluation of commonly used techniques for transposing data across 167 catchment pairs. This large-scale evaluation is unique within the UK. The best method identified for transposing gauged flow data was based on normalising the gauged flow time series by an estimate of catchment mean flow, derived from a regional statistical model, and re-scaling the resultant time series by an estimate of mean flow for the ungauged catchment. The evaluation studies confirmed current thinking that a good analogue catchment is one that is close, similar in terms of climatic regime and that is hydrogeologically similar - and thus has a similar response to rainfall inputs. Furthermore, the analysis has demonstrated that transposing of data from upstream or downstream (connected) analogues, generally, results in a very good simulation of the times series structure of the flow data at the target site and an acceptable simulation of mean flow. This contrasts with the case where the analogue catchment is from an adjacent, unconnected catchment. In this case the simulations obtained are generally poor.

A comparison with the regionalised form of MODB demonstrated that if the analogue catchment is connected to the ungauged catchment, and meets the selection criteria specified, the transposing of gauged flow data will generally result in a better simulation of stream flow. The time series fit is generally better, however the use of the regionalised rainfall runoff model will generally give a marginally better water balance. In cases where the analogue and ungauged catchments are not connected, the use of a regionalised rainfall runoff model will generally give a better simulation of stream flow than the use of data from the analogue catchment. In this context, the regionalised model represents a significant advance over the current practice of transposing gauged data from analogue catchments.

## 10.2 FURTHER WORK

Ideas for further research are subdivided into two areas.

*Input data and model structure*

The research has identified that accurate climatic input data and an ability to identify a contributing catchment area are crucial to the process of closing an adequate water balance. Hence these issues are crucial to identifying loss module parameters that are a function of the catchment type rather than being an artefact of errors in the data/and or the assumption of a closed water balance. Further research into the relationships between these errors and the model structure is required. This will to give guidance on how the problem may be ameliorated by changes to the model structure and/or identifying critical deficiencies within the input data.

Analysis of the model behaviour within different catchment types identified that in some catchment types parts of the model structure were redundant. The simple model MODB, in some respects, represented the best compromise between minimising redundancy whilst maintaining conceptual realism. Research on how the optimal model structure might be defined as a function of the catchment type could prove to be extremely fruitful, particularly in the extremes of groundwater catchments and impermeable, upland catchments.

As discussed in Chapter 1, the study specifically did not address snow storage. This is not a major issue as snow storage is only important within some highland catchments within the north of the country. However, it is an area that should be addressed for completeness.

The majority of catchments used within the study are small catchments. If the models were to be applied within large basins a strategy for subdividing the basin up into appropriate sub catchments needs to be developed. The issue of channel routing of stream flows from these sub-catchments would also need to be addressed within this strategy.

*Calibration and regionalisation*

Within this study, model calibration and the subsequent regionalisation of parameters were treated as distinct areas of research. The biggest criticism of the multivariate regression based model for predicting parameters is that there is inevitably a degree of structural dependency between the parameters for a particular parameter vector; i.e. they are covariant to some degree.

It may be feasible to address this by merging the model calibration and regionalisation schemes. In this approach, the relationships between model parameters and catchment characteristics could be sequentially optimised with the order based on the strength of the relationships with catchment characteristics. Between each sequential step, the model would be re-calibrated using the predicted values for those parameters that had been regionalised. This type of approach is currently being evaluated within the context of the regionalisation of continuous simulation models for predicting extreme floods at CEH-Wallingford (Calver, *pers comms*).

The time series objective functions used in the study have addressed the accuracy of the simulated flows within individual time steps. It is suggested that the utility of fuzzy functions that address the entire time series should be evaluated. An example of this type of function could be based on the mean and variance of the distribution of the percentage differences between observed and simulated flows across the entire stream flow. Calibration would involve the minimisation of these statistics.

# References

Abdulla, F.A. and Lettenmaier, D.P. 1997. Application of regional parameter estimation schemes to simulate the water balance of a large continental river. Journal of Hydrology. **197**, 258-285.

Abdulla, F.A. and Lettenmaier, D.P. 1997. Development of regional parameter estimation equations for a macroscale hydrologic model. Journal of Hydrology. **197**, 230-257.

Agung, B., Ibrahim, A.B. and Cordery, I. 1995. Estimation of recharge and runoff volumes from ungauged catchments in eastern Australia. Hydrological Sciences Journal. **40**, 499-516.

Allen, R.G., Smith, M., Pereira, L.S., and Perrier, A. 1994. An update for the calculation of reference evapotranspiration. ICID Bulletin, **43**, 35-92.

Andersson, L. and Harding, R.J. 1991. Soil-moisture deficit simulations with models of varying complexity for forest and grassland sites in Sweden and the UK. Water Resources Management, **5** , 25-46.

Arnell, N.W. and King, R. 1997. The impacts of climate change on water resources. In DETR/The Met. Office. Climate Change and it's impacts: a Global Perspective. Department of the Environment, Transport and the Regions. pp 10-11.

Arnell, N.W. and Reynard, N.S. 1996. The effects of climate change due to global warming on river flows in Great Britain. Journal of Hydrology. **183**, 397-424.

Arnell, N.W. 1999. A simple water balance model for the simulation of streamflow over a large geographic domain. Journal of Hydrology. **217**, 314-335.

Aston, A.R. 1979. Rainfall interception by eight small trees. Journal of Hydrology, **42**, 383-396.

Bayliss, A. 1999. Catchment Descriptors. Flood Estimation Handbook Vol.5. NERC, Wallingford.

Bell, F.G. 1985. *Engineering Properties of Soils and Rocks*. Butterworths, London.

Bergström, S. and Forsam, A. 1973. Development of a conceptual deterministic rainfall-runoff model. Nordic Hydrology . **4**, 141-170.

Beven, K and Binley, A.M. 1992. The Future of Distributed Models: Model Calibration and Uncertainty Prediction. Hydrological Processes, **6**, 279-298.

Beven, K. 1979. A sensitivity analysis of the Penman-Monteith actual evapotranspiration estimates. Journal of Hydrology, **44**, 169-190.

Beven, K. 1989. Changing Ideas in Hydrology: The Case of Physically-Based Models Journal of Hydrology. **105**, 157-172.

Beven, K. 1993. Prophecy, Reality and Uncertainty in Distributed Hydrological Modelling. Advances in Water Resources. **16**, 41-51.

Binley, A.M., Beven, K.J., Calver, A. and Watts, L.G. 1991.Changing Responses in Hydrology: Assessing the Uncertainty in Physically Based Model Predictions. Water Resources Research. **27**, 1253-1261.

Blackie, J.R. and Eeles, C.W.O. 1985. Lumped Catchment models. In Hydrologcial Forcasting (edited by M.G. Anderson and T.P. Burt). John Wiley and Sons Ltd.

Boorman, D.B., Hollis, J.M. and Lilly, A. 1995. Hydrology of Soil types: a hydrologically-based classification of the soils of the United Kingdom. Institute of Hydrology Report No. 126. Wallingford UK.

Boughton, W.C. 1966. A mathematical model for relating runoff to rainfall with daily data. Civil Eng. Trans., I.E. Aust. **8** , 83-93.

British Standards Institute. 1996. BS7843 ,Guide to the acquisition and management of meteorological precipitation data. Section 2.4 Areal Rainfall.

Brookes, R.H. and Corey, A.T. 1971. Hydraulic properties of porous media. Colarado State University, Hydrology Paper No. 3.

Burn, D.H. and Goel, N.K. 2000. The formation of groups for regional flood frequency estimation. Hydrological Sciences Journal. **45** , 97–112.

Burn, D.H., Boorman, D.B. 1993. Estimation of Hydrological Parameters at Ungauged Catchments Journal of Hydrology. **143**, 429-454.

Burnash, R.J.C. 1995. The NWS River Forecast System-Catchment modelling. Chapter 10, Computer Models of Watershed Hydrology. Editor V.P. Singh, Water Resources Publications, Colorado, USA.

Calder, I.R. 1990. Evaporation in the uplands. Chichester. John Wiley & Sons, London. UK.

Calder, I.R. 1986a. The influence of land use on water yield in upland areas of the UK. Journal of Hydrology, **88** ,201-211.

Calder, I.R. 1986b. A stochastic model of rainfall interception. Journal of Hydrology, **89**, 65-71.

Calder, I.R. and Newson, M.D. 1979. Land-use and upland water resources in Britain - a strategic outlook. Water Resources Bulletin, **15** , 1628-1639.

Calder, I.R., Harding, R.J., and Rosier, P.T.W. 1983. An objective assessment of soil-moisture deficit models. Journal of Hydrology, **60**, 329-355.

Calver, A. 1996. Development and experience of the TATE rainfall runoff model. Proc. Inst. Civ. Eng. Water Marit. Energy. **118**, 168 –176.

Calver, A., Lamb, R. and Morris, S.E. 1999. River Flood Frequency Estimation using continuous simulation runoff modelling. Proc. Inst. Civ. Eng. Water Marit. Energy. **136**, 225 – 234.

Chiew, F.H.S. and McMahon, T.A. 1991. Improved modelling of the groundwater processes in MODHYDROLOG. Proc. Hydrol. And Water Resources Symp., Perth, Western Australia, October 1991.

Chiew, F.H.S., Stewardson, M.J. and McMahon T.A. 1993. Comparison of six rainfall-runoff modelling approaches. Journal of Hydrology. **144**, 1-36.

Crawford, N.H. and Linsley, R.K. 1966. Digital Simulation on Hydrology: Stanford Watershed Model IV. Stanford University Technical Report No. 39. Stanford University, Palo Alto.CA.

Cox , D.R. and Hinkley, D.V. 1974. Theoretical Statistics. Chapman and Hall, London.

Duan, Q., Gupta, V.K. and Soorooshian, S. 1993. A shuffled complex evolution approach for effective and efficient global minimization. Journal of Optimisation Theory and Applications. **74**, 501-521.

Duan, Q., Soorooshian, S. and Gupta, V.K. 1992. Effective and efficient global optimization for conceptual rainfall-runoff model. Water Resources Research, **28**, 1015-1031.

Edijanto, N.O. and Michel, C. 1989. Un modele pluie –debit journalier a trois parameteres. La Houille Blanche, **2**, 113-122.

Egbuniwe, N. and Todd, D.K. 1976. Application of the Stanford Watershed Model to Nigerian Watersheds Water Resources Bulletin. **12**, 449-460.

Faures, J-M, Goodrich, D.C., Woolhiser, D.A. and Sorooshian, S. 1995. Impact of small-scale spatial rainfall variability on runoff modeling. Journal of Hydrology **173**, 309-326.

Fawthrop N. 1992. Naturalisation of the Orton River Flow Record. National Rivers Authority. Anglian Region. Unpublished report.

Franchini, M. and Pacciani, M. 1991. Comparative Analysis of Several Conceptual Rainfall-Runoff Models. Journal of Hydrology, **122**, 161-219.

Franchini, M., Galeati, G. and Berra S. 1998. Global optimization techniques for the calibration of conceptual rainfall-runoff models. Hydrological Sciences Journal. **43**, 443-458.

Fuller, R.M., Groom, G.B. and Jones, A.R. 1994. The Land Cover Map of Great Britain: An Automated Classification of Landsat Thematic Mapper Data. Photogrammetric Engineering and Remote Sensing, **60**, 553-561.

Gash, J.H.C. 1979. An analytical model of rainfall interception by forests. Quarterly Journal of the Royal Meteorological Society, **105**, 43-55.

Greenfield, B. 1984. The Thames Catchment Model. Thames Water Authority. Unpublished Report.

Grindley, J. 1970. Estimation and mapping of evaporation. IAHS Publication No. **92**, 200-213.

Gustard, A, Bullock, A. and Dixon, J.M. 1992. Institute of Hydrology Report 108. Low Flow Estimation in the United Kingdom.

Gustard, A. and Sutcliffe, M.F. 1986. Low flow study of Northern Ireland. Report to Department of the Environment, Northern Ireland.

Gustard, A., Marshall, D.C.W. and Sutcliffe, M.F. 1987. Low flow estimation in Scotland. IH Report No. 101, Wallingford, UK.

Hall, J.K. and Nott, M.R. 1994. The naturalisation of flow records by decomposition. British Hydrological Society National Meeting on Flow Naturalisation Using Hydrological Models, London 17th March 1994.

Hall, R.L. and Harding, R. 1993. The water use of the Balquhidder catchments : a processes approach. Journal of Hydrology, **145**, 285-314.

Harding, R., Hall, R.L., Neal, C., Roberts, J.M., Rosier, P.T.W., and Kinniburgh, D.G. 1992. Hydrological impacts of broadleaf woodlands: implications for water use and water quality. National Rivers Authority, Bristol.

Hollis, J.M. and Woods, S.M. 1989. The measurement and estimation of saturated soil hydraulic conductivity, SSLRC report to MAFF.

Hollis, J.M. 1989. A methodology for predicting Soil Wetness Class from soil and site properties. SSLRC report to MAFF.

Holmes M.G., Young A.R. 2000a. Estimation of Mean Flow in the UK. Low Flows Studies 2000. Technical Report 11. NERC, Wallingford. *In press.*

Holmes M.G., Young A.R. 2000a, 2000b. Estimation of Flow Duration Curves in the UK. Low Flows Studies 2000. Technical Report 12. NERC, Wallingford. *In press.*

Horton, R.E. 1938. The interpretation and application of runoff plot experiments with reference to soul erosion problems. Soil Science Society Am., Proc., **3**, 340-249.

Hough, M. 1996. The Meteorological Office Rainfall and Evaporation Calculation System: MORECS Version 2.0 (1995). An update to Hydrological Memorandum 45. The Meteorological Office, UK.

Houghton-Carr H.A. 1999. Assessment criteria for simple conceptual daily rainfall-runoff models. Hydrological Sciences Journal, **44**, 237-262.

Hughes, D.A. and Sami K. 1994. A semi-distributed, variable time interval model of catchment hydrology – structure and parameter estimation procedures. Journal of Hydrology , **155**, 265-291.

Hughes, D.A. and Sami, K. 1995. A semi-distrbuted, variable time interval model of catchment hydrology – structure and parameter estimation procedures. Journal of Hydrology **155**, 265-291.

Hughes, D.A. 1989. Estimation of the parameters of an isolated event conceptual model from physical catchment characteristics. Hydrological Sciences Journal**. 34**, 63-78.

Hughes, D.A. 1995. Monthly rainfall-runoff models applied to arid and semiarid catchments for water resource estimation purposes. Hydrological Sciences Journal. **40**, 751-769.

Hughes, D.A. and Smakhtin, V. 1996. Hydrological Sciences Journal. **41**, 851-871.

Hyoms, C.M.1980. Development of a Soil Moisture Model for the Estimation of Percolation. Unpublished M.Phil thesis, The City University.

Ibbit, R.P. 1970. Systematic parameter fitting for conceptual models of catchment hydrology, unpublished Ph.D. Thesis, Imperial College of Science and Technology, University of London, London, United Kingdom.

Institute of Geological Sciences 1977, Hydrogeological maps of England Wales and Scotland.

Institute of Hydrology. 1985. The FSR rainfall-runoff model parameter estimation equations updated. Flood Studies Supplimentary Rep. No. 16. Institute of Hydrology, Wallingford.

Jakeman, A.J. and Hornberger, G.M.,1993. How much complexity is warranted in a rainfall runoff model ? Water Resources Research, **29**, 2637-2649.

Jakeman, A.J., Littlewood, I.G., Whitehead, P.G. 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. Journal of Hydrology, **117**, 275-300.

Jarboe, J.E. and Haan, C.T. 1974. Calibrating a water yield model for small ungaged watersheds. Water Resources Research. **10**, 256-262.

Johnston, PR and Pilgrim, DH. 1976. Parameter optimization for watershed models. Water Resources Research. **12**, 477-486.

Jolley, T.J.and Wheater, H.S. 1997a. An investigation into the effect of spatial scale on the performance of a one-dimensional water balance model. Hydrological. Processes. **11**, 1927-1944.

Jolley, T.J.and Wheater, H.S. 1997b. The introduction of runoff routing into large-scale hydrological models. Hydrological. Processes. **11**, 1917-1926.

Jones, S.B. 1983. The estimation of catchment average point rainfall profiles. Institute of Hydrology, UK, Report No. 87.

Lamb, R., Beven, K. 1997. Using interactive recession curve analysis to specify a general catchment storage model, Hydrology and Earth System Sciences, **1**, 101-113.

Leavesly, G.H. and Stannard, L.G. 1995. The Precipitation-Runoff Modelling Stsyem-PRMS. Chapter 9, Computer Models of Watershed Hydrology. Editor V.P. Singh, Water Resources Publications, Colorado, USA.

Lighthill, M.J. and Witham, G.B. 1955. On kinematic waves. 1: Flood movement in long rivers. Proc. R. Soc., Series A, **229**, 281-316.

Littlewood, I.G. and Jakeman, A.J. 1994. A new method of rainfall runoff modelling and its applications in catchment hydrology. In P.Zanetti (ed) Environmental Modelling (Volume II) Computational Mechanics Publications, Southampton, UK, 141-171.

Littlewood, I.G. and Parker, J. 1997. PC-IHACRES User Guide. Institute of Hydrology, Wallingford, UK.

Magette, W.L., Shanholtz, V.O. and Carr, J.C. 1976. Estimating Selected Parameters for the Kentucky Watershed Model From Watershed Characteristics. Water resources Research. **12**, 472-476.

Manley, R.E. 1977. The soil moisture component of mathematical catchment simulation models., Journal of Hydrology. **35**, 341-356.

Manley, R.E. 1978. The use of a hydrological model in water resources planning. Poc.Instn Civ. Engrs, Part 2, 1978, **65**, 223-235.

Manley, R.E. 1992a. HYSIM Technical Guide, unpublished software documentation.

Manley, R.E. 1992a. HYSIM User Guide, unpublished software documentation.

Manley, R.E. 1978. Simulation of Flows in Ungaged Basins Hydrological Sciences Bulletin. **23**, 85-101.

Manly, B.F.J. 1992. The Design and Analysis of Research Studies. Cambridge University Press, Cambridge, UK.

Matheron, G. (1963) Principles of geostatics, Economic Geology, **58**, 1246-1266.

Meteorological Office 1981. Handbook of meteorological Instruments, Vol.5. HMSO, London. 2nd edition.

Midgley, D.C., Pitman W.V and Middleton B.J. 1994. Surface Water Resources of South Africa, Users Manual. Water Resources Commission, WRC Report No. 298/1/94.

Monteith, J.L. 1965. Evaporation and environment. 19th Symposia of the Society for Experimental Biology. Canbridge, Cambridge University Press, **19**, 205-234.

Moore, R.J. 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrological Sciences Journal, **30**, 273-297.

Moore, R.J., Austin R.M. and Carrington D.S. 1994. Evaluation of FRONTIERS and Local Radar Rainfall Forecasts for Use in Flood Forecasting Models. R&D Note 225.National Rivers Authority, UK.

Moore, R.J., Jones D.A. and Black K.B. 1989. Risk assessment and drought management in the Thames basin. Hydrological Sciences Journal, **34**, 705-717.

Moore, R.J. 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrological Science Journal, **30**, 273-297.

Morris, D. and Flavin, R. 1990. A digital terrain model for hydrology. Proc. 4th International Symposium on Spatial Data Handling, Zurich, 1, 250-262.

Morris, D. and Heerdegen, R. 1988. Automatically derived catchment boundaries and channel networks and their hydrological applications. Geomorphology, **1**, 131-141.

Murdock, R.U. and Gulliver, J.S. 1993. Prediction of River Discharge at Ungaged Sites with Analysis of Uncertainty, Journal of Water Resources Planning and Management. **119**, 473-487.

Nandakumar, N. and Mein, R.G. 1997. Uncertainty in rainfall-runoff model simulations and the implications for predicting the hydrological effects of land-use change. Journal of Hydrology, **192**, 211-232.

Nash, J.E. and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models, part 1: a discussion of principles. Journal of Hydrology, **10**, 282-290.

Natural Environment Research Council (NERC) 1975. Flood Studies Report. Wallingford, UK.

Natural Environment Research Council (NERC) 1980. Low Flow Studies report. Wallingford, UK.

Nelder, J.A. and Mead, R. 1965. A simplex method for function minimization. Computer Journal, **7**, 308-313.

Nielson, S.A. and Hansen, E. 1973. Numerical simulation of the rainfall-runoff process on a daily basis. Nordic Hydrology. **4**, 171-190.

Obled, Ch, Wendling, J. and Beven, K. 1994. The sensitivity of hydrological models to spatial rainfall patterns: an evaluation using observed data. Journal of Hydrology, **159**, 305-333.

Paturel, J.E., Servat E., and Vassiliadis A. 1995. Sensitivity of conceptual rainfall-runoff algorithms to errors in input data- case of the GR2M model. Journal of Hydrology, **168**, 111-125.

Penman, H.L. 1948. Natural evaporation from open water, bare soil and grass. Proceedings of the Royal Society, A. **193**, 120-145.

Penman, H.L. 1949. The dependence of transpiration on weather and soil conditions. Journal of Soil Science, **1** , 74-89.

Penman, H.L. 1963. Vegetation and hydrology. Technical Communication 53, Commonwealth Bureau of Soils, Harpenden.

Philip, J.R. 1957. The theory of infiltration, I The infiltration equation and its solution. Soil Sci., **83**, 345-357.

Pilling, C and Jones, J.A. 1999. High resolution climate change scenarios: implications for British runoff. Hydrological Processes. **13**, 2877-2895.

Pirt, J and Bramley, E.A. 1985. The application of simple moisture accounting models to ungauged catchments. J Inst Water Eng Sci. **39**, 169-177.

Pitman, W.V. and Kakebeeke, J.P. 1991. The Pitman model – into the 1990s. Proceedings of the 5th South African national Hydrology Symposium, Stellenbosch, November 1991.

Post, D.A and Jakeman, AJ. 1996. Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments. Hydrological Processes. **10**, 877-892.

Quick, M.C. 1995. The UBC watershed model. Chapter 8, Computer Models of Watershed Hydrology. Editor V.P. Singh, Water Resources Publications, Colorado, USA.

Robson, A. and Reed, D.W. 1999. Statistical Procedures for Flow Frequency Estimation. Flood Estimation Handbook Vol.3. NERC, Wallingford.

Rodda, J.C. 1967, The systematic error in rainfall measurement. J. Inst. Water Eng., **21**, 173-177.

Rosenbrock, H.H. 1960. "An Automatic method of finding the greatest or least value of a function. Computer Journal, **3**, 175-184.

Rutter, A.J., Kershaw, K.A., Robins, P.C., and Morton, A.J. 1971. A predictive model of rainfall interception in forests, I. Derivation of the model from observations in a plantation of Corsican pine. Agricultural Meteorology, **9**, 367-384.

Sefton, C.E.M and Boorman, D.B. 1997. A regional investigation of climate change impacts on UK streamflows. Journal of Hydrology. **195**, 26-44.

Servat, E. and Dezetter, A. 1991. Selection of calibration objective functions in the context of rainfall runoff modelling in a sudanese savannah area. Hydrological Sciences Journal. **36**, 307-330.

Shaw, E.M. 1988. Hydrology in Practice – 2nd Edition. Van Nostrand Reinhold (Interantional), London, UK.

Sibson, R. 1982. A brief description of natural neighbour interpolation. In: Barnett V. (ed), Interpreting Multivariate data , 21-36. John Wiley & Sons.

Singh, V.P. 1995. Watershed Modelling. Chapter 1, Computer Models of Watershed Hydrology. Editor V.P. Singh, Water Resources Publications, Colorado, USA.

Soorooshian S. and Gupta V.K. 1983. Automatic calibration of conceptual rainfall-runoff models: the question of parameter observability and uniqueness. Water Resources Research. **19**. 251-259.

Sorooshian Sand Gupta. 1995. Model calibration. Chapter 2, Computer Models of Watershed Hydrology. Editor V.P. Singh, Water Resources Publications, Colorado, USA.

Sorooshian, S and Gupta, VK. 1985. The analysis of structural identifiability: theory and application to conceptual rainfall-runoff models. Water Resources Research. **21**, pp 487-495.

Southern Water Authority 1979. Report on low flow studies for the Southern Water Authority area. Unpublished report.

Spackman, E. 1993. Calculation and Mapping of Rainfall Averages for 1961-90. British Hydrology Society Meeting on Areal Rainfall, 15th Dec. 1993, University of Salford, Manchester. Unpublished paper.

Speers, D.D. 1995. SSARR Model. Chapter 11, Computer Models of Watershed Hydrology. Editor V.P. Singh, Water Resources Publications, Colorado, USA.

Storm, B., Jensen, K.H. and Refsgaard, J.C. 1988. Estimation of Catchment Rainfall Uncertainty and its Influence on Runoff Prediction. Nordic Hydrology, **19**, 77-88.

Sugawara, M. 1995. The Tank Model. Chapter 6, Computer Models of Watershed Hydrology. Editor V.P. Singh, Water Resources Publications, Colorado, USA.

Systat, 1998 SYSTAT 8.0 for Windows: Statistics.SPSS Inc. Chicago.

Thian Yew Gan, Dlamini E.M & Biftu G.F., 1997. Effects of model complexity and structure, data quality, and objective functions on hydrological modeling. Journal of .Hydrology. **192**, 81-103.

Thiessen, A.H. 1911. Precipitation for large areas. Monthly Weather Review, **1,** 1082-1084.

Todini, E. 1996. The Arno rainfall-runoff model., Journal of .Hydrology. **175**, 339-382.

Todini, E. 1988. Rainfall-Runoff modelling: past, present and future. Journal of Hydrology, **100**, 341-352.

Tulu, T. 1991. Simulation of streamflows for ungauged catchments. Journal of Hydrology. **129**, 3-17.

Vogel, R.M. and Sankarasubramanian, A. 2000. Spatial scaling properties of annual stream flow in the United States. Hydrological Sciences Journal. **45**, 465-487.

Wallace, J.S., Roberts, J.M. and Roberts, A.M. 1982. Evaporation from Heather Moorland in North Yorkshire, England. *Proc. Symp. Hydrol.* Research Basins Sonderb, Landeshydrologie, Bern.

Watts, G. 1994. Naturalisation of the Babingley Flows. Internal Report. National Rivers Authority – Anglian Region. Unpublished report.

Weeks, W.D. and Hebbert, R.H.B. 1980. A comparison of Rainfall-Runoff Models. Nordic Hydrology. **11**, 7-24.

Wilby, R.L. 1994a. Modelling the Relative Impact of Weather, Land use and Groundwater Abstraction on Low Flows. R&D Note 268.National Rivers Authority. UK.

Wilby, R., Greenfield, B. and Glenny, C. 1994b. A coupled synoptic-hydrological model for climate change impact assessment. Journal of Hydrology. **153**, 265-290.

WMO. 1974. Inter-comparison of Conceptual Models used in Operational Hydrological Forecasting. Geneva, Report 7. WMO Geneva, Switzerland.

Wood, E.F., Lettenmaier, D.P. and Zartarian, V.G. 1992. A land-surfave hydrology paramterisation with subgrid variability for general circulation models. J. Geophys. Res. **97**, 2717-2728.

WRc. 1990. A Resource Model of the Great Ouse River System. Report No. CO 2504-M.

Wright, I.R. and Harding, R. 1993. Evaporation from natural mountain grassland. Journal of Hydrology, **145**, 267-283.

Yapo, P.O., Gupta, V.H. and Sorooshian, S. 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology. **181**, 23-48.

Young, A.R., Gustard, A., Bullock, A., Sekulin, A.E., Croker, K.M. 2000. A river network based hydrological model for predicting natural and influenced flow statistics at ungauged sites, Science of the Total Environment. **251/252**, 293-304.

Young, A.R.. Round, C.E., Gustard A. 2000. Spatial and Temporal variations in the occurrence of low flow events in the UK. Hydrology & Earth System Sciences. **4**, 35-45.

Young, A.R. and Sekulin, A.E. 1996. Naturalised River Flow Records of the Essex Region: Phase III Final Report. Institute of Hydrology unpublished Client Report.

Zhao, R.J., Zhuang, Y., Fang L.R., Lin X.R. and Zhang Q.S. 1980. The Xinanjiang model. In: hydrological forecasting (Proc. Oxford Symp., April 1980) IAHS Publ. No. 129.

# Appendix A:

# Model evaluation graphs for the

# Babingley Brook at Castle Rising

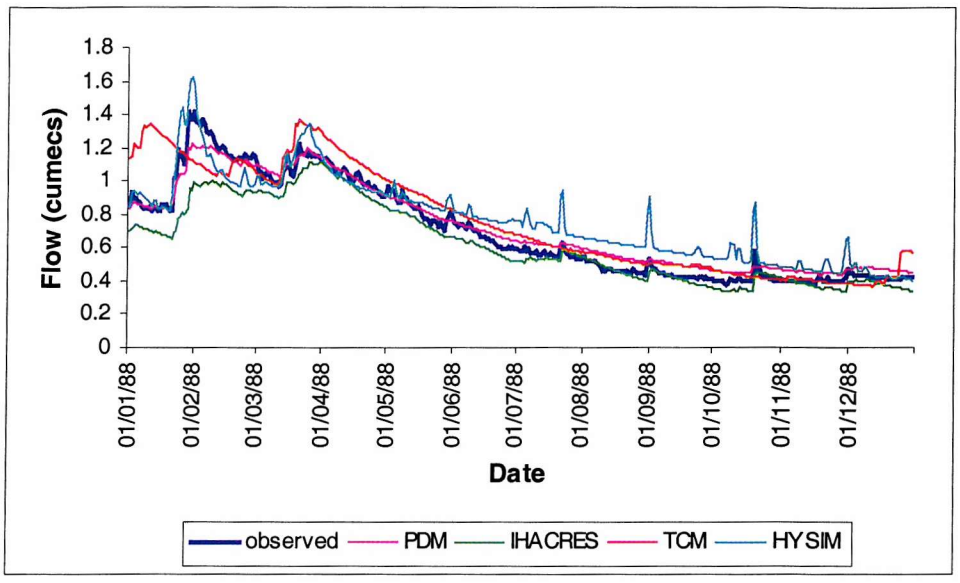



**Figure A.1: Calibration period: observed and simulated hydrographs for 1988**

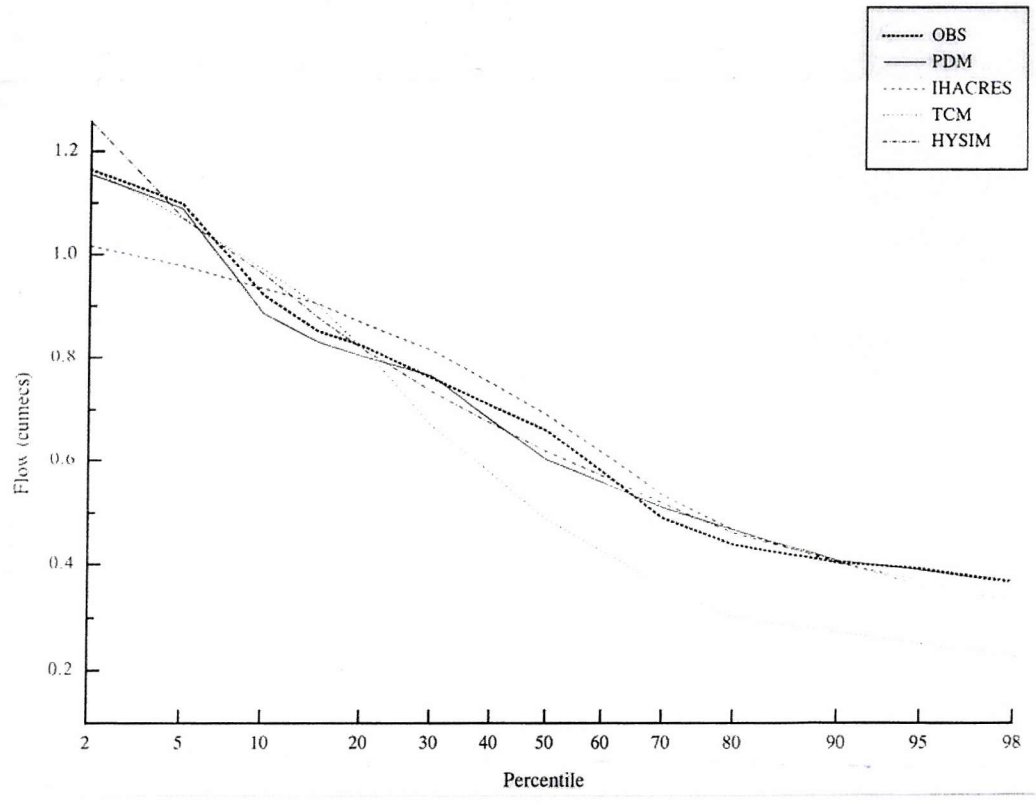



**Figure A.2: Calibration period: simulated and observed flow duration curves**

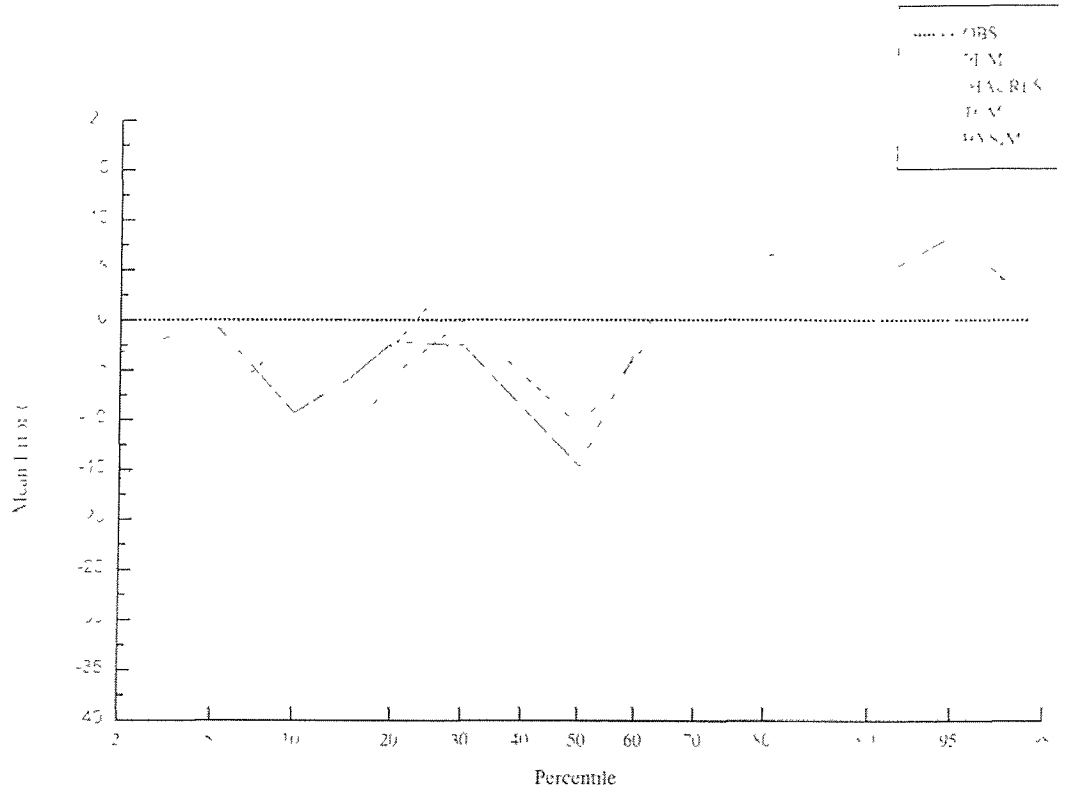



**Figure A.3: Calibration period: mean simulation errors at observed percentile points**

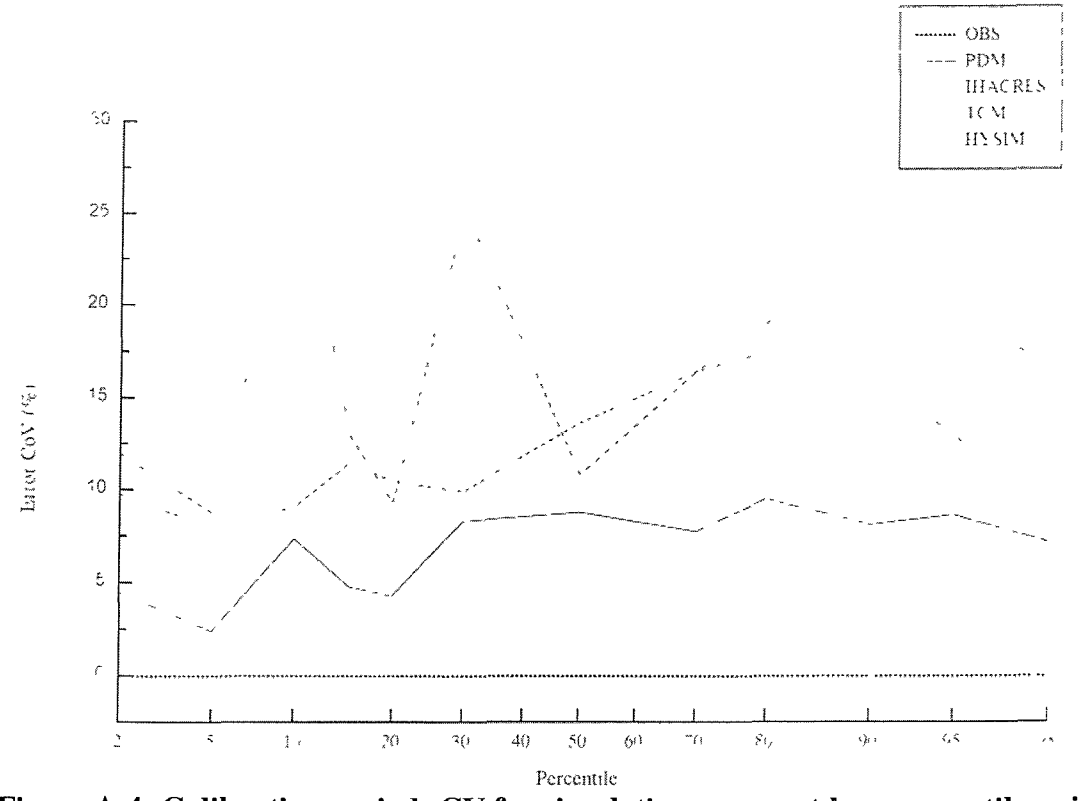



**Figure A.4: Calibration period: CV for simulation errors at key percentile points**

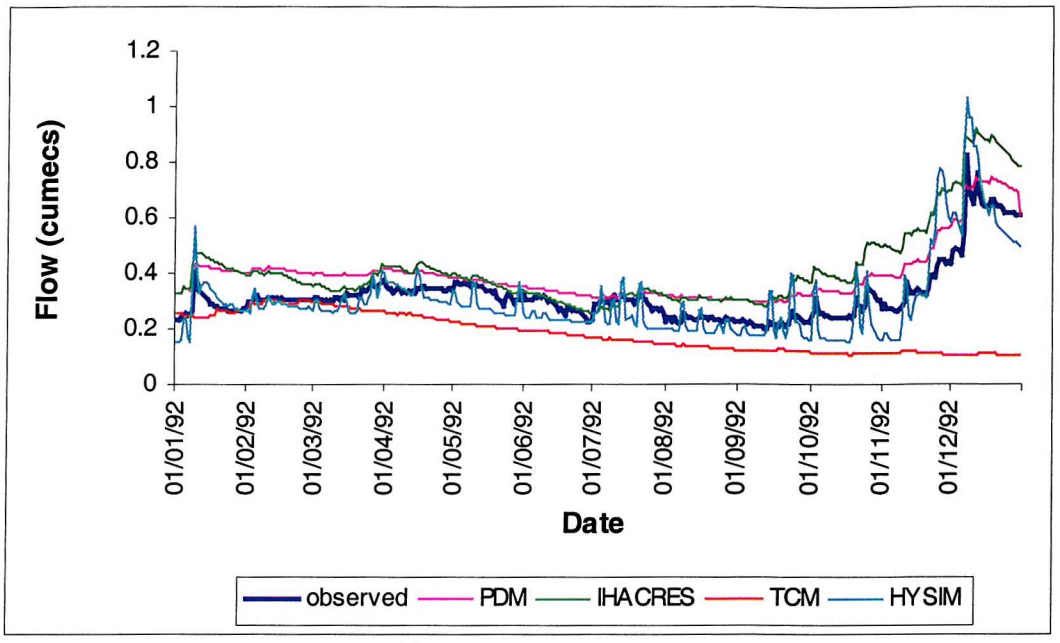



**Figure A.5: Evaluation period: observed and simulated hydrographs for 1976**

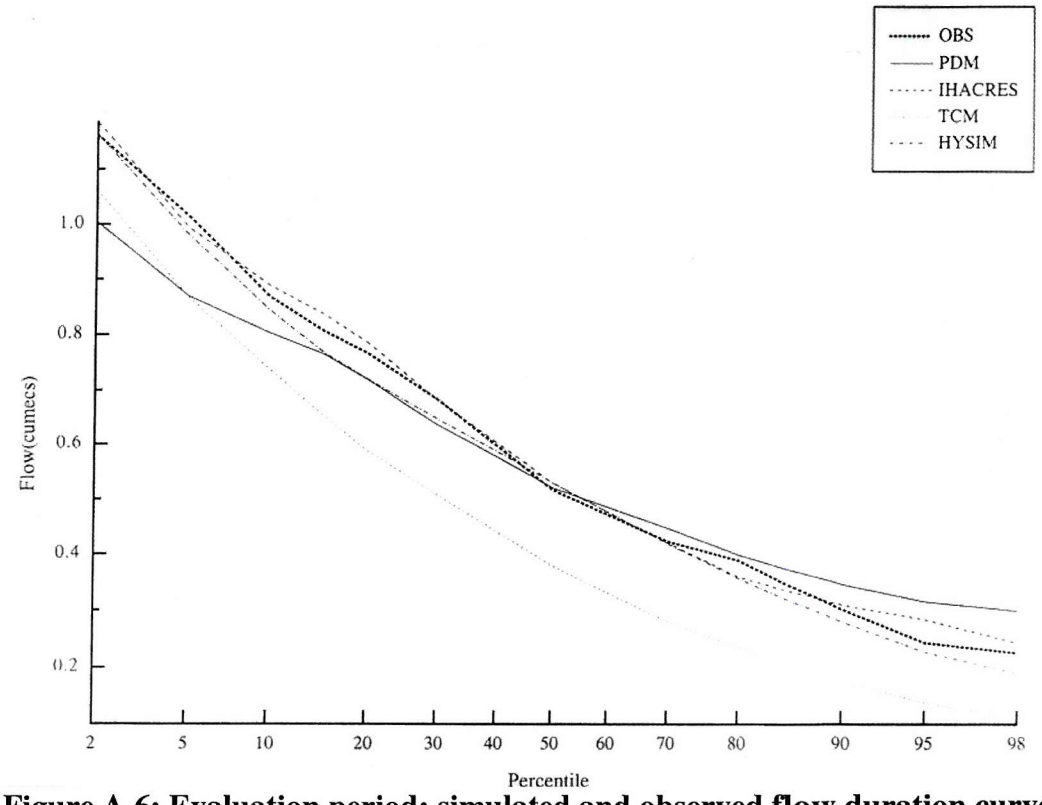



**Figure A.6: Evaluation period: simulated and observed flow duration curves**

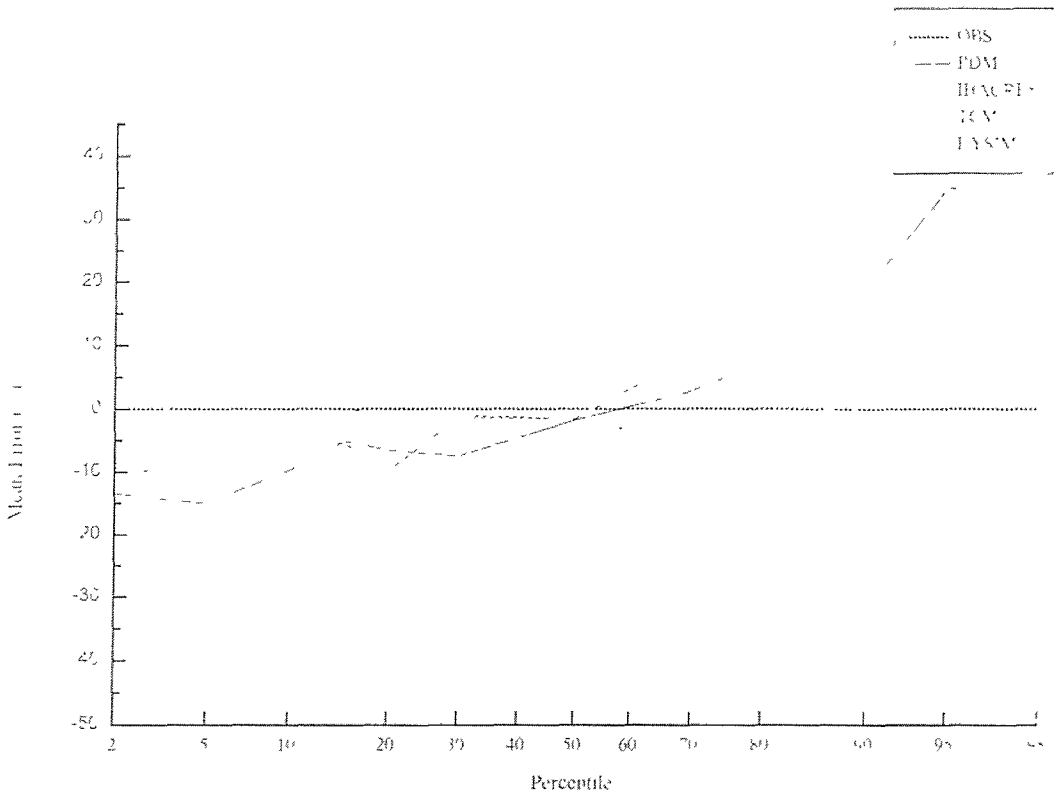



**Figure A.7: Evaluation period: mean simulation errors at observed percentile points**

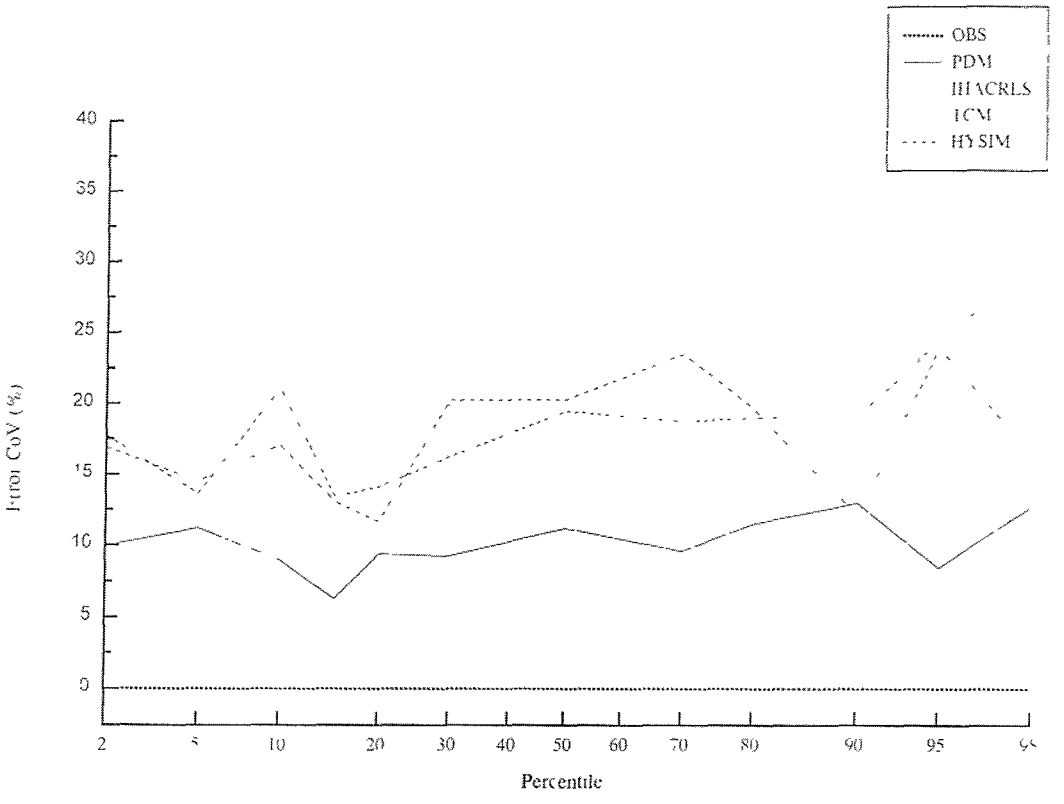



**Figure A.8: Evaluation period: CV for simulation errors at key percentile points**

# Appendix B:
# The model calibration scheme and the estimation of catchment specific parameter limits

# B The model calibration scheme and the estimation of catchment specific parameter limits

## B.1 THE CALIBRATION SCHEME

The model calibration scheme developed for the research is a three-stage approach based upon a Monte Carlo sampling strategy of the feasible parameter space. The scheme initially identifies a realistic parameter space prior to identifying parameter vectors that result in equally acceptable simulations of stream flow over the calibration period. A feasible parameter space is one in which the parameter ranges are suitable for all catchments, and is defined by upper and lower limits for parameter values. The realistic parameter space is defined as a catchment specific subset of the feasible parameter space that contains the significant areas of attraction within the objective function surface(s).

- In the first stage the loss module is calibrated to identify a feasible parameter space for the loss module, starting from initial upper and lower feasible limits for the individual parameters.
- In the second stage realistic optimal, or target values for the LF_OBJ and EFF objective functions are estimated.
- In the third stage these target values are then used to identify 300 equally likely parameter vectors for the model given the accuracy of the input data and the requirement of the simulated flows to give adequate values for the four objective functions, defined within chapter 6.

The following sub-sections describe the three stages, and the application of each stage to a catchment. In practice the whole procedure was fully automated.

### B.1.1 First stage: The loss module

The objective functions for this stage are model BIAS and error in BFI predicted by the loss module. The procedure for calculating these statistics is described in Section 6.2 of Chapter 6.

There are up to 10 loops within the loss module calibration stage. Within each loop, parameter vectors for the loss module are randomly selected from the feasible parameter space, assuming a uniform distribution for each parameter between upper and lower limits, this is termed a shot. The model is then run over the calibration period for the shot parameter vector, and the BIAS and simulated BFI are calculated for the simulation run. This sampling of the parameter space is repeated until 3000 valid parameter vectors are obtained. A shot is valid if the modelled output from the loss module meets threshold values for BIAS and DBFI. Once the defined number of valid shots has been obtained, the thresholds for BIAS, DBFI and feasible parameter limits are reduced and the next loop initiated. This procedure is repeated until either the end point criteria for BIAS and DBFI are reached or 10 loops have been completed.

The starting point values for DBFI and BIAS are that the a simulation shot must have a BIAS equal to, or less than 40% and that DBFI must be less than 0.2. The initial end point criteria were a BIAS of ±3% and a DBFI of 0.15. At the end of each loop the parameter limits are reset to the maximum and minimum values observed within the valid shots within the loop and the objective function criteria are set to the mean values of the criteria over the valid shots. The total number of shots required to identify the specified number of valid shots depends on the loop number. To obtain 3000 valid shots for loop typically required a total of between 100,000 and 500,000 shots for the loop. This large number of shots ensured that a thorough search of the loss model parameter space is made within each loop, such that a reduced parameter space containing the significant areas of attraction within the objective function surfaces is obtained.

In application to a catchment, if the end point criteria were obtained prior to the $10^{th}$ loop the number of loops was truncated. If the end point criteria were not met within the 10 loops, the end point criteria for BIAS was relaxed to ±5%. The catchment was rejected from the second stage of the calibration procedure on the basis of being unable to close an acceptable water balance if this relaxed end point BIAS criterion could not be met given the DBFI constraints.

The choice of 3000 valid shots and the endpoint criteria are subjective. The choice of valid shots provides for a good sampling of the parameter space with a loop without incurring

excessive computational effort. The end point criteria for DBFI of 0.15 is a reflection of the empirical nature of the calculation of the observed BFI, and the lack of direct correspondence with the partitioning of effective rainfall by the model. What the use of this criterion does ensure however, is that the model will be capable of simulating the broad hydrological response characteristics of the observed regime (as represented by the observed BFI) using the final "realistic" parameter space for the loss module defined within this stage.

The end point criterion of a BIAS of 3% was selected after discussions with staff of the National River Flow Archive at the Centre for Ecology and Hydrology - Wallingford. The figure represents an estimate for what a good quality, well maintained gauging station structure should be able to achieve in terms of hydrometric accuracy, at a confidence level of approximately one standard deviation. If anything, this criterion is stringent. The relaxation of the BIAS criterion to ±5%, if the initial end-point criterion of ±3% was not attainable, was introduced to boost the sample size and reflects the tight constraint that the first end point places upon the model. Assuming a gaussian measurement error distribution for gauged flow, this is tighter than a 95% confidence interval, which would be approximately ±6%. In practice, in the majority of catchments where the initial end-point criterion could not be met the relaxed criterion also could not be met. This is indicative of either a gross error in the climatic input data or a violation of the assumption of a closed water balance. The latter error may be due to an incorrectly defined contributing area or gauge bypass, both of which can occur in groundwater fed catchments.

Once the realistic parameter space for the loss module parameters has been identified within stage 1, the parameter limits describing this space feed into the second and third stages of the calibration procedure.

### B.1.2 Second Stage: Identification of realistic limits for the LF_OBJ and EFF objective functions.

The objective within the second stage of calibration is to identify realistic, acceptable values for the LF_OBJ and EFF objective functions. Within the second stage the fit of the full model is optimised, constrained by the realistic parameter space for the loss module and a feasible parameter space for the two parameters of the routing module. The second stage is again based on a Monte-Carlo sampling strategy, although this time there is no end point criteria specified for the objective functions. The four objective functions are all used within this second stage. The target objective function limits for BIAS and DBFI are the end point values from the loss module calibration stage. These limits are not modified during this second stage, or the third stage of calibration.

In application, the initial target objective function limits for LF_OBJ and EFF are set to 5000 (i.e. a large value) and zero respectively. The endpoint values for these latter objective functions can not be specified, other than as the perfect fit equating to zero and one respectively. Within this second stage, shot samples are repeatedly selected from the parameter space, and for each shot that meet the target values for BIAS and DBFI, the values of LF_OBJ and EFF are calculated. A shot is accepted as valid if both of the objective function values represent an improvement over current target values for the LF_OBJ and EFF functions. If a shot is valid, the target values for the LF_OBJ and EFF functions are reduced by an amount equivalent to 20% of the improvement observed, thus gradually tightening the target limits for these objective functions for subsequent shots. The process is repeated until:

- either further appreciable improvement in the objective function limits can not be found (identified as four successive valid shots in which the maximum improvement for either objective function was less than 5% of the current target value for the function),
- or until the total number of model runs exceeds 250,000.

Through trials this second stage was found to successfully identify the approximate region within the parameter space of the minima for each objective function, given the constraints of input data accuracy and conflicting regime requirements from the LF_OBJ and EFF

functions. The gradual tightening of the objective function limits prevents the excessive minimisation of one or other of the objective functions at the expense of being able to minimise the other objective function.

### B.1.3 Third stage: Identification of equally valid parameter vectors for the model

The objective of the third stage in the scheme is to identify 300 valid shots using the optimal target values of the four objective functions identified within the first and second stages. For this third stage the feasible parameter space is repeatedly sampled until 300 valid shots are identified. A fuzzy logic approach is adopted for testing whether a particular shot is valid. A shot is accepted as being valid if:

- the shot values of BIAS and DBFI are less than or equal to the end point criteria from the loss module,
- AND the value of one of the routing objective functions (LF_OBJ or EFF) is equal to or better than the reasonable target limit identified within the second stage,
- AND the value of the second objective function is within +20% of the identified reasonable target limit for the second objective function.

The setting of a fuzzy limit allows for a trade off between the quality of fit over the whole time series and that at low flows. This is an attempt to mimic the type of trade off a modeller will make when evaluating the quality of a model fit through visual inspection. The 300 parameter values obtained in this way can be considered as being equally likely-given the various requirements for accurate simulation, as represented by the objective functions and the inaccuracies within the input data.

This heuristic, Monte–Carlo based calibration procedure differs from conventional search algorithms in that the selection of individual valid parameter vectors are independent of one another. The approach is computationally intensive despite the use of the first stage to reduce computational overheads. It does mean however, that the parameter space is well sampled to provide a range of valid parameter vectors from which to make the final selection of a parameter vector for use within the parameter regionalisation studies, described in chapter 7. The final choice of parameter vector is discussed within Section 6.6

of Chapter 6.

### B.1.4 Limitations of the calibration scheme

The approach is very computationally intensive, a typical runtime for the calibration scheme, calibrating the model over a 10year period of observed daily mean flows, was in the order of three hours, or approximately eight catchment data sets a day. This was for a FORTRAN implementation of both the model and calibration scheme running on a 400mHz Intel Pentium II based PC. There is a direct trade off between the size of the feasible parameter space, the required number of valid shots and the length of time the code took to run.

The method requires a definition of parameter space for each parameter (discussed in the subsequent section) and the definition of a distribution for the parameter from which to draw samples. This involves a degree of subjectivity. As will be discussed, the setting of a feasible limit for Kg in MODA was difficult. To address this the limits for Kg were programmed in as "fuzzy" limits. During the first three iteration loops for the loss module the feasible limits for Kg are not fixed. If a valid shot has a value of Kg that occurs within 5% of either limit the space for Kg is moved by 10% in that direction. After loop three it was empirically demonstrated on a wide range of catchment types that the scheme had found the feasible space containing the significant areas of attraction within the Kg parameter space. The same approach was adopted for the initial setting of feasible limits for the time constant for the quick-flow routing reservoir, K1.

A problem that was not addressed was that the approach assumes that parameters are independent of one another. This is not the case, however it is primarily is an issue of sampling in-efficiency, the impact of which will again be reduced by the iterative reduction of the realistic parameter space in Stage 1.

## B.2 THE ESTIMATION OF CATCHMENT SPECIFIC PARAMETER SPACES

### B.2.1 A priori estimation of the interception model depth parameter for MODA

Early calibration trials identified a significant problem in identifying realistic values of the interception model parameter $\gamma$ within MODA, as a result of large covariance between $\gamma$ and the other parameters within the loss module: Cmax, b and Kg. An example of this is presented in Figure B.1 for the Bure at Ingworth (34003) (a small, dry, permeable catchment with geology comprised of sands, gravels and loams). In this figure, the loss model parameters from 3000 valid shots for this catchment are plotted as a function of $\gamma$. It can be seen that Kg is function of the value of $\gamma$. Furthermore, high values of b do not occur at low values of $\gamma$ and, conversely, low values of Cmax do not occur at low values of $\gamma$. This covariance is a consequence of the requirement for the calibration scheme to close an adequate catchment water balance.

The interception store only plays a role when there is a significant Soil Moisture Deficit (SMD) within the model soil moisture store, that results in a reduction of the evaporation rate to below the potential rate (Equations 4.14 and 4.23, Chapter 4). When the SMD is not significant, evaporation takes place from the soil stores at the potential rate; this is the same rate at which evaporation takes place from the interception store. During periods with a significant SMD, the interception store prevents the premature alleviation of the SMD, within the soil moisture store by intercepting incident precipitation that would otherwise reduce the SMD and hence increase the drainage from the store. As $\gamma$ increases, resulting in a greater interception capacity, a increasing percentage of the precipitation incident on the catchment surface will be intercepted, and evaporated at the potential rate rather than evaporated from the soil moisture store at a rate governed by the soil moisture status of that store.

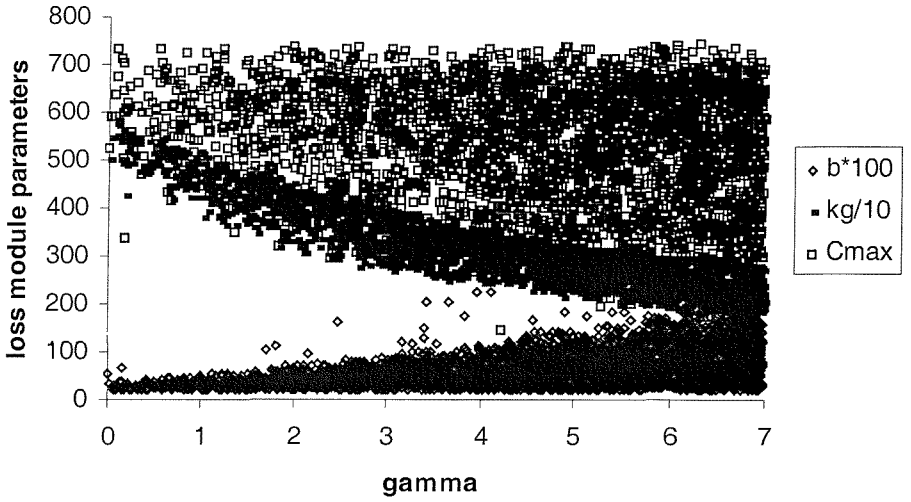



**Figure B.1 An example of loss module parameter covariance for catchment 34003.**

This can be compensated for within the Pareto based loss module by either increasing the value of b or reducing the value of Cmax, which in turn will reduce $\overline{C}$, and hence Smax (Equation 4.14, Chapter 4). As $\overline{C}$ is reduced the percentage of time that significant soil moisture deficits exist will increase, hence the degree of covariance observed between $\gamma$ and these two parameters. If Kg is low then the drainage from the soil moisture store for a given soil moisture content, S(t), is high (Equation 4.18, Chapter 4), thus a low value of Kg will enhance the development of a significant SMD, also counteracting the impact of a high interception storage capacity.

It is not surprising that the interception parameter is not that identifiable. It is only significant when there are significant soil moisture deficit, which represent a relatively small percentage of the time, particularly in the wetter areas of the UK. However the interception model is important within the conceptual structure for MODA for drought conditions *A priori* values for $\gamma$ were estimated for each catchment based on the vegetation classes within the catchment and the work, described in Section 4.2.1of Chapter 4, on estimating annual interception losses for specific vegetation types. The interception capacity of a vegetation class is a function of the ability of the vegetation to catch incident precipitation, which is related to size and density of the plant. The fraction of the precipitation that can be intercepted is also controlled by the efficiency with which water can be evaporated from the vegetation surface. The latter is related to aerodynamic

roughness, which is in turn related to size and vegetation structure. Based on knowledge of the structure of the vegetation within each class of the ITE land cover classification class and the regional behaviour of the $\gamma$ parameter described in Section 4.2.1 of Chapter 4, an *a priori* value for $\gamma$ was defined for each vegetation class. These values are summarised in Table B.1 A catchment value for $\gamma$ was estimated for each catchment using

$$\gamma_{catchment} = \frac{\sum A_i \gamma_i}{\sum A_i} \qquad \textbf{(B.1)}$$

where $\gamma_i$ is the maximum daily interception loss for vegetation class, i, and $A_i$ is the fractional coverage of vegetation type, i.

The fractional extents of vegetation classes within each catchment were obtained from the ITE land cover classification coverage. The limitations of this approach are:

- the values will not necessarily be optimal;
- it is assumed that the ITE land cover classes are representative of the historical land cover in the catchment and that the fractional extents for individual classes are fully covered by the class vegetation type.

These limitations are offset by the minor role that the interception module plays in determining the overall fit of the model over a wide range of climatic conditions.

**Table B.1** ***A priori*** **values of the interception parameter, γ, for the ITE land classes**

| Land Class | γ (mm) | Land Class | γ(mm) |
|---|---|---|---|
| 0 Unknown | 0 | 13 Lowland Heather | 3 |
| 1 Sea | 0 | 14 Scrub / Orchard | 3 |
| 2 Water (inland) | 0 | 15 Deciduous | 3 |
| 3 Beach | 0 | 16 Coniferous | 5.9 |
| 4 Saltmarsh | 0 | 17 upland bog | 0.5 |
| 5 Lowland Grass / Heather | 1.5 | 18 Arable | 2.25 |
| 6 Pasture | 1 | 19 Ruderal Weed | 0.5 |
| 7 Meadow | 1 | 20 Suburb | 0.5 |
| 8 Rough Grass | 1.5 | 21 Urban | 0 |
| 9 Montane Grass | 1.5 | 22 Bare | 0 |
| 10 Grass / Heather Mix | 2 | 23 Felled Forest | 4.45 |
| 11 Heather | 3 | 24 Lowland Bog | 0.5 |
| 12 Bracken | 3 | 25 Open Shrub Heath | 2.25 |

## B.3 THE ESTIMATION OF A CATCHMENT SPECIFIC PARAMETER SPACE FOR THE SLOW FLOW RESERVOIR TIME CONSTANT

In identifying limits for the slow flow reservoir time constant, Kb, the objective was to determine a reduced realistic parameter space to improve the efficiency of the third stage of the calibration procedure. As discussed in Section 4.3 of Chapter 4, a linear reservoir formulation was used for both routing reservoirs. The Institute of Hydrology (1980) describes an automated method for identifying a time constant for an exponential recession function (equivalent to a linear reservoir with zero inflows), based on the extraction and plotting of all recession curves from observed stream flow data using this transformation. Drawing a line parallel to the enveloping line where individual recessions run together can be used to derive an estimate of the recession constant. Lamb and Beven (1997) describe a method for using master recession curve analysis to help in defining the form of an unknown underlying storage function, rather than for an assumed storage function.

In the approach developed for this study, specific limits for Kb were determined for each catchment by identifying and estimating the time constants from individual gauged recession periods for a catchment. To restrict the recession periods to when, conceptually, the total flow is dominated by the outflow from the slow flow reservoir, only recession periods below a threshold flow of Q80 were considered. Below a Q80 flow it can be assumed that catchment stream flow is dominated by the release of water from storage

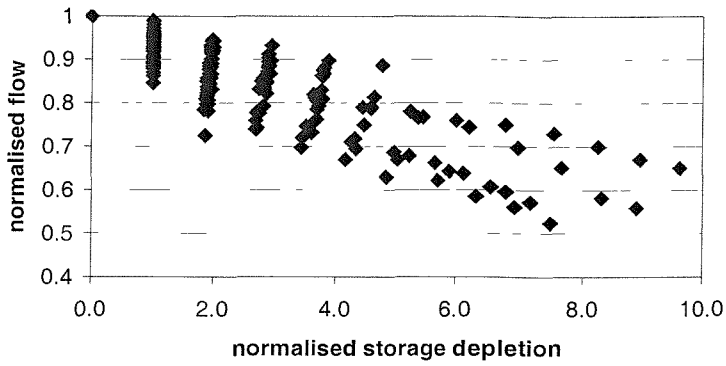
within the catchment. A period of recession for this study was defined as starting when the flow on day (i+1) is less than the flow on day (i). The impact of the choice of threshold was evaluated by considering the recession behaviour in ten percentile increments between Q50 and Q90 for both permeable and impermeable catchments. It was found that the number of significant recession periods, meeting the acceptance criteria, decreased rapidly for flows higher than the Q80 flow. The end point of the n day recession was taken as either when the flow on day (i+n) is greater than the flow on day (i+n-1) or when the incident precipitation on day (i+n) was greater than two percent of the flow on day (i+n). The latter constraint was used to minimise the impact of the assumption that the inflow into the conceptual reservoir is zero. A further constraint of three days was placed on the minimum length of the recession period. For each catchment, individual periods of recession were identified from the calibration period flow data. Under zero inflows, and with a daily time step, the outflow from the reservoir within a time step can be approximated using Equation 4.24 (Chapter 4). A plot of the cumulative flow from the start of recession to day (i+n-1) represents the volume by which the reservoir is depleted. The gradient of a straight line fitted to this plot will be an estimate of the reciprocal of the time constant Kb. An estimate of Kb was obtained using a least squares fit for each of the m recession periods identified for a catchment. For each catchment, the mean values of Kb and the corresponding standard error of the mean were calculated from the sample of m estimates of Kb. Using the central limit theorem, a 95% confidence interval was constructed for the sample mean as

$$\overline{\hat{K}}b - 1.96 \text{x} \text{S.E} \leq Kb \leq \overline{\hat{K}}b + 1.96 \text{x} \text{S.E}, \qquad \textbf{(B.2)}$$

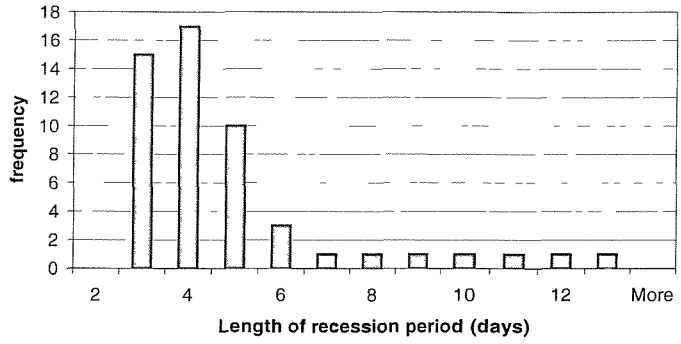
where S.E is the standard error of the estimated mean value. This procedure is demonstrated in Figure B.2 for catchment 67010. This catchment is one of four used to illustrate parameter identifiability issues within Section 6.5 of Chapter 6. The recession curve analysis identified 56 individual recession periods over the calibration period for this catchment. The flow – storage depletion relationships are plotted as points in Figure B.2(a). These points are normalised by the flow at the start of the recession period to facilitate plotting on the same axis. A histogram of the length of recession periods indicates that the modal length of a recession period was four days with, as would be expected, a greater incidence of short recession periods. A histogram of the estimated values of Kb for each

recession period is presented in Figure B.2(c) and the relationship between Kb and length of recession period is presented in Figure B.2(d). Figure B.2(d) illustrates that there is a much greater variation in the estimate of Kb for short recession periods, primarily because these estimates are generated from a smaller number of points. There may also be a quick flow recession component, and/or the assumption of zero inflows into the reservoir may be violated. The longer recession periods generally yield estimates of Kb in the region of 400 hrs. The mean value of Kb and the standard error of the mean are 447 and 25 days respectively. The 95% confidence interval used to define the upper and lower limits of the parameter space for Kb was hence [397,497].

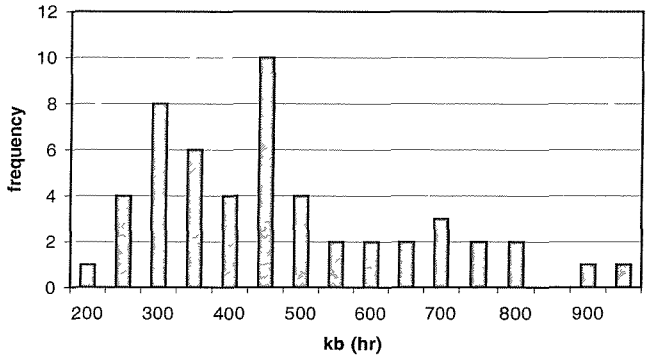
For 13 stations this analysis yielded fewer than 10 recession periods, which was regarded as being an insufficient number for this analysis. These stations all had BFI estimates of less than 0.37 indicating that they were all fairly flashy. A simple regression model was constructed for predicting Kb from BFI estimates using the mean Kb and BFI estimates for the remaining 226 catchments. The relationship is presented in Figure B.3. The model explained 58% of the variance ($R^2$) and had a factorial standard error of 66%. The upper and lower parameter limits for the 13 stations were estimated using the 95% predictive confidence interval from this model.



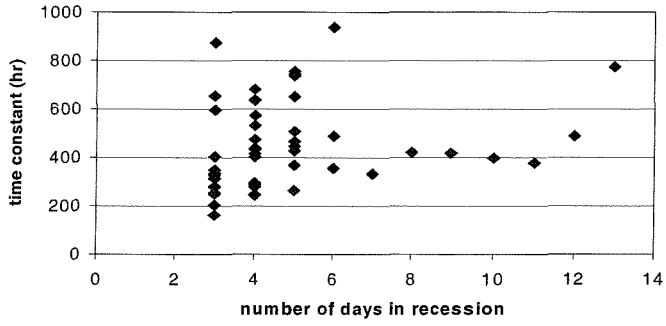

a) Flow Depletion plots




b) Histogram of recession periods




c) Histogram of estimated Kb values




d) Estimated values of Kb as a function of recession period

**Figure B.2 The recession curve analysis as applied to catchment 67010.**

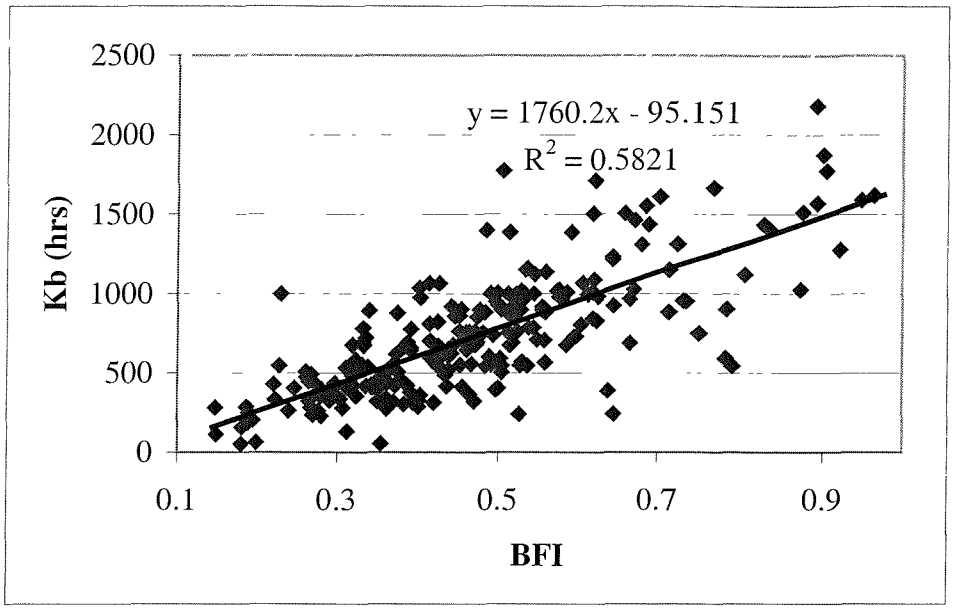



**Figure B.3 Relationship between Kb and BFI for 229 catchments**

# Appendix C:
# An evaluation of the behaviour of MODA and MODB within four case study catchments

# C An evaluation of the behaviour of MODA and MODB within four case study catchments

## C.1 INTRODUCTION

As discussed within Chapter 6, the approach to the calibration of the models for this study recognises that uniqueness of solution is unlikely to exist, given the uncertainty associated with the input data and the sometimes-conflicting simulation requirements of different objective functions. The result is that there are ranges of stream flow simulations (with associated objective function values) that are equally as likely realisations of the optimal simulation. In this appendix the results from the calibration scheme for four catchments are used to illustrate the dependencies of the models on catchment type, the potential parameter covariance issues for MODA and the impact of errors in the input data upon the loss module parameters. The behaviour of MODB within the same catchments is also discussed within the appendix. The four case study catchment selected for the discussion represent the following climatological and hydrological permutations:

- dry- permeable :- The Bure at Ingworth (34003);
- dry- impermeable:- The Pinn at Uxbridge (39098);
- wet- permeable:- Sydling Water at Sydling St. Nicholas (44006);
- wet- impermeable:- The Gelyn at Cynefail (67010).

These four catchments are discussed in detail within Chapter 6.

## C.2 AN EVALUATION OF MODA MODEL FITS WITHIN THE FOUR CASE STUDY CATCHMENTS.

Range statistics regarding the parameters, the objective functions and goodness of fit ranges for the valid shots identified for each of the catchments are summarised in Table C.1 and Table C.2 for the catchments considered. Table C.1 presents the 95, 50 and 5 percentile values (derived non-parametrically) for the individual variables. The variance within the range for each variable is presented within Table C.2.

The variance is summarised non-parametrically using:

$$VAR = \frac{V95 - V5}{2V50} \qquad \textbf{(C.1)}$$

This measure is analogous to the more familiar coefficient of variation from gaussian statistics. The behaviour of MODA within each catchment is discussed in more detail within the following sub-sections.

**Table C.1 Percentile values for parameter, objective function and goodness of fit ranges**

| | 34003 | | | 39098 | | | 44006 | | | 67010 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **P95** | **P50** | **P5** | **P95** | **P50** | **P5** | **P95** | **P50** | **P5** | **P95** | **P50** | **P5** |
| **Cmax** | 794 | 599 | 273 | 787 | 532 | 139 | 543 | 441 | 374 | 140 | 109 | 72 |
| **b** | 1.06 | 0.79 | 0.43 | 2.11 | 1.37 | 0.15 | 0.16 | 0.06 | 0.00 | 3.92 | 3.11 | 1.78 |
| **K1** | 139 | 84 | 56 | 53 | 26 | 12 | 411 | 222 | 114 | 45 | 30 | 16 |
| **Kb** | 1779 | 1430 | 1047 | 416 | 250 | 94 | 1719 | 803 | 498 | 496 | 452 | 401 |
| **Kg** | 6661 | 5593 | 4122 | 104267 | 62879 | 26581 | 8969 | 6760 | 5624 | 893 | 803 | 752 |
| **LF_OBJ** | 30 | 27 | 25 | 83 | 73 | 62 | 32 | 27 | 22 | 57 | 48 | 41 |
| **BIAS** | 3 | 1 | -3 | 3 | 0 | -3 | 3 | 0 | -3 | -3 | -3 | -3 |
| **EFF** | 0.67 | 0.62 | 0.51 | 0.74 | 0.68 | 0.56 | 0.76 | 0.64 | 0.60 | 0.77 | 0.76 | 0.71 |
| **BEQ95** | 10 | -20 | -33 | 41 | -29 | -60 | 36 | 26 | -3 | 3 | -7 | -15 |

**Table C.2 Variance statistics for parameter, objective function and goodness of fit ranges**

| | **34003** | **39098** | **44006** | **67010** |
|---|---|---|---|---|
| Cmax | 43 | 61 | 19 | 32 |
| b | 40 | 71 | 143 | 35 |
| Kg | 49 | 79 | 67 | 49 |
| K1 | 26 | 64 | 76 | 11 |
| Kb | 23 | 62 | 25 | 9 |
| LF_OBJ | 9 | 14 | 17 | 16 |
| BIAS | 3 | 3 | 3 | 0 |
| EFF | 13 | 26 | 32 | 4 |
| BEQ95 | 26 | 71 | 16 | 9 |

### C.2.1 The Bure at Ingeworth, 34003

A Pearson correlation matrix presenting significant correlations between variables of greater than 0.4 at a significance level of 95% is presented in Table C.3 for catchment 34003. Scatter plot matrices for the correlation between parameters, and between the parameters and objective functions/goodness of fit measures are presented in Figure C.1a

and Figure C.1b. The histograms for parameter values (Figure C.1a) show that parameters within the middle of the range occur more frequently thus confirming that the choice of parameters was not constrained by the restriction on the feasible parameter space. The objective functions presented in Figure C.1b are BIAS, EFF and LF_OBJ, and an additional objective function, the Euclidean Objective Function, EOF. The use of the EOF is discussed in the Chapter 6, Section 6.6. The BEQ95 goodness of fit measure is also included within the second scatter plot.

There is a high positive correlation between the loss module parameters: Cmax, b and Kg. The range of the variation, with VAR values between 22 and 44, is high (Table C.3). The routing parameter, K1 and Kb do not correlate with each other, or the loss module parameters. The range of Kg values is comparable with that for the other permeable catchment, 44006 (Table C.1).

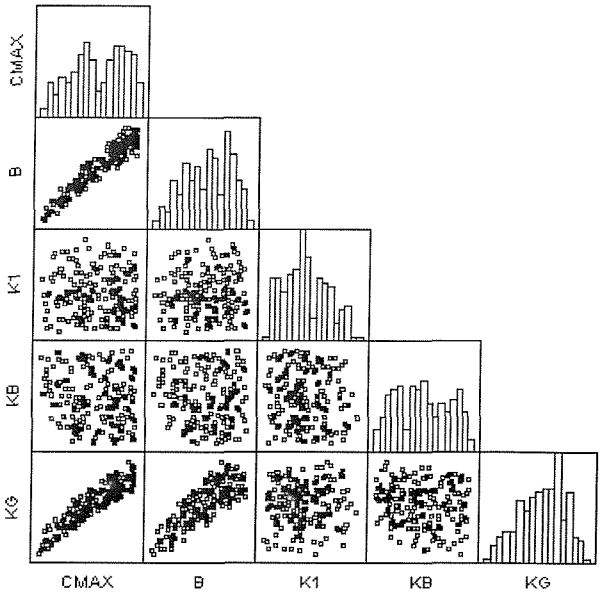
**Table C.3 Correlations for 34003**

| | CMAX | b | K1 | Kb | Kg | LF_OBJ | BIAS | EFF | BEQ95 |
|---|---|---|---|---|---|---|---|---|---|
| **CMAX** | 1.00 | | | | | | | | |
| **B** | 0.94 | 1.00 | | | | | | | |
| **K1** | | | 1.00 | | | | | | |
| **Kb** | | | | 1.00 | | | | | |
| **Kg** | 0.92 | 0.79 | | | 1.00 | | | | |
| **LF_OBJ** | | | -0.55 | -0.64 | -0.58 | 1.00 | | | |
| **BIAS** | | | | | | | 1.00 | | |
| **EFF** | | | -0.51 | | | 0.46 | | | |
| **BEQ95** | 0.56 | 0.52 | | 0.74 | | -0.86 | | | 1.00 |

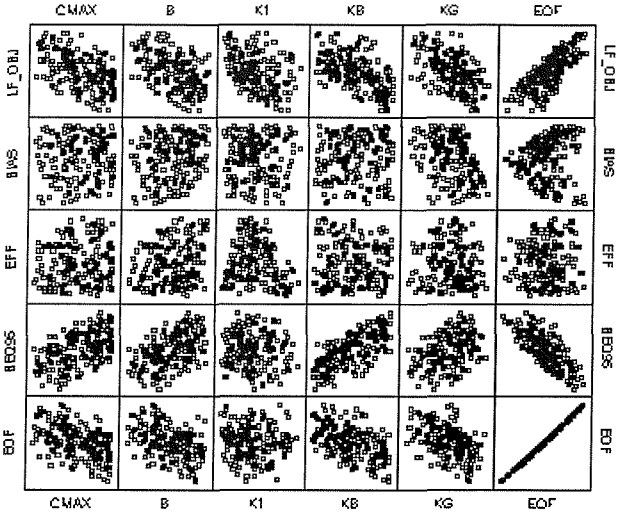
LF_OBJ is inversely correlated with both routing parameters, indicating that both of the routing reservoirs are important in controlling the fit of the model over the lowest third of the flow distribution. The LF_OBJ objective function improves (becomes smaller) as the routing reservoirs become less responsive (K1 and Kb become larger).

BIAS is not strongly correlated with any of the parameters, but BEQ95 is strongly correlated with the loss module parameters. Evaporative losses are dependent upon the soil moisture deficit, this is expressed as the depth of water held in storage across the catchment as a fraction of the total depth of water that can be held in storage (Equation 4.23, Chapter 4). For a given set of inputs, the ability of the model to generate a significant soil moisture deficit (one that limits the evaporation rate) is controlled by the value of Smax. Smax is

equal to the mean storage capacity over the catchment, $\overline{C}$, which in turn is controlled by the values of Cmax and b (Equation 4.14, Chapter 4).




(a)




(b)

**Figure C.1 Scatter plots for catchment 34003.**

A large value for $\overline{C}$ will lead to an over prediction of losses and hence an under prediction of mean flow as significant SMD will not occur as frequently. A smaller value will lead to an under prediction of losses as significant SMD will frequently build up. As discussed in Chapter 6, significant SMDs do build up in this catchment, as a consequence of the low rainfall and thus the loss module is important in controlling the BIAS (error at mean flow) of a model fit.

The parameters Cmax and b compensate for each other in controlling $\overline{C}$, and hence losses, which explains the strong covariance between these parameters in this case. The mean depth, $\overline{C}$, is in the order of 200 mm at the 5% value of Cmax and 400 mm at the 95% value of Cmax. This would indicate that significant SMD are more likely to exist at low values of Cmax and b. The rate at which water drains from the soil store is inversely proportional to the parameter Kg. Higher values of Cmax yield correspondingly higher values of $\overline{C}$, and hence Smax. The runoff for a given depth of rainfall is correspondingly lower. This is compensated for by an increase in Kg, which will in turn limit drainage from the base of the store, hence generating more direct runoff from the store.

The EFF objective function is negatively correlated with the quick flow time constant, K1. EFF is a measure of the overall time series model fit and is biased towards the fit at higher flows. The EFF objective function improves with a decreasing value for K1 whereas the LF_OBJ function improves with an increasing value of K1. There is therefore a trade off with K1 between the fit at low flows and the overall fit of the hydrograph. This is confirmed by the weak positive correlation between EFF and LF_OBJ.

The bias at Q95 flow, BEQ95, is negatively correlated with LF_OBJ. However, it is also positively correlated with b and Cmax, indicating that the under estimation of BEQ95 is less at high values of Cmax and b (as the model consistently has a negative bias at Q95). The reason for this is not obvious; but it is probably associated with the partitioning of effective rainfall between the quick and slow flow routing paths. BEQ95 is also positively correlated with Kb, indicating that the under estimation is lower for larger values of Kb and hence a lower gradient recession curve. The model was unable to obtain a better fit at Q95 flows within this catchment because of the trade off between LF_OBJ and EFF.

### C.2.2 Pinn at Uxbridge (39098)

The correlation matrix for this catchment is presented in Table C.4 and the scatter plot matrices in Figure C.2. Significant SMD occur in this catchment for about 40% of the time. Within this catchment there is an even stronger degree of co-linearity between Cmax and b than that observed with 34003, but the correlation between Kg and Cmax and b is not significant. The range for both parameters encompasses most of their feasible parameter space, indicating that these parameters are not readily identifiable. The values of Kg are

very high, resulting in very little drainage to the slow flow routing reservoir, and again the values cover quite a wide range. The range of BIAS indicates that the model can easily close a water balance. The negligible base flow (observed BFI =0.09) within this catchment explains the lack of correlation between Kg and the other loss module parameters, as there is negligible drainage through the base of the soil moisture store to the slow flow routing reservoir. This also explains why Kg is not readily identifiable. The main restriction on the acceptable values of Kg is only that they have to be large, thus resulting in negligible drainage.

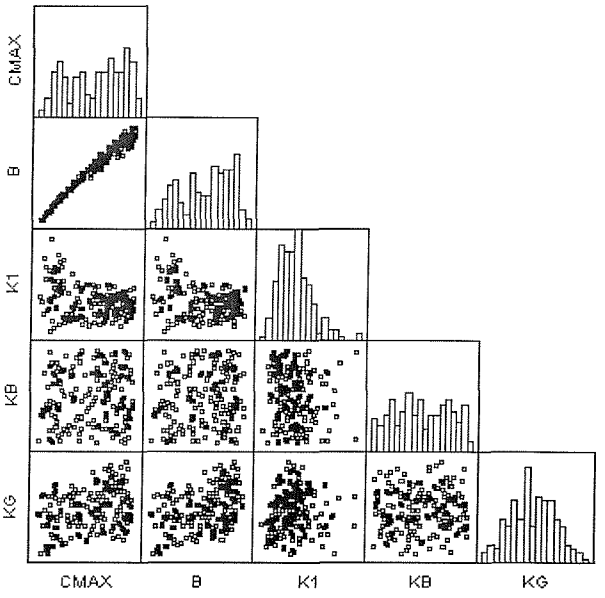
**Table C.4 Correlation matrix for 39098**

| | CMAX | b | K1 | Kb | Kg | LF_OBJ | BIAS | EFF | BEQ95 |
|---|---|---|---|---|---|---|---|---|---|
| **CMAX** | 1.00 | | | | | | | | |
| **B** | 0.98 | 1.00 | | | | | | | |
| **K1** | | | 1.00 | | | | | | |
| **Kb** | | | | 1.00 | | | | | |
| **Kg** | | | | | 1.00 | | | | |
| **LF_OBJ** | 0.45 | 0.57 | | | 0.85 | 1.00 | | | |
| **BIAS** | | | | | | | 1.00 | | |
| **EFF** | | | 0.52 | | 0.53 | 0.44 | | 1.00 | |
| **BEQ95** | | | | | 0.72 | | | 0.44 | 1.00 |

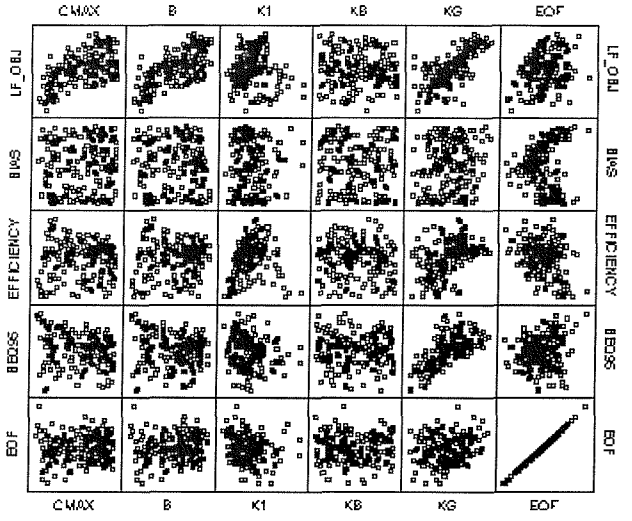
The range and magnitudes of values for K1 and Kb are commensurate with the values observed for the other impermeable catchment 67010 (Table C.1). Interestingly, considering the range of values within the valid shots for K1 and Kb, these time constants do not show a strong correlation with the LF_OBJ and EFF objective functions. This implies that the small-scale variations in objective function values are not sensitive to magnitude of the variation observed in K1 and Kb. The catchment has a very low base flow. The average volume of water passing through the slow flow routing path was 9% of the total volume of water and thus the slow flow routing path, and hence Kb, is not critical in the fit of the model within this catchment.

The LF_OBJ function correlates positively with all of the loss model parameters, particularly Kg, and thus the way in which effective rainfall is partitioned within the loss module. These positive correlations imply that better fits are obtained for smaller values of Kg, Cmax and b. Given the small base flow component, this in turn implies that the direct runoff response (in part controlled by b) to smaller rainfall events is important in controlling LF_OBJ.

EFF is positively correlated with Kg, K1 and LF_OBJ implying a trade off again between the overall hydrograph fit and the fit at low flows. BEQ95 is positively correlated with Kg and thus the model will over estimate Q95 flows for larger values of Kg and under estimate for smaller values of Kg. This behaviour is not intuitively correct and reflects that the slow flow component is a negligible component of the hydrograph fit and that the model fit is controlled by the quick flow routing component. The role of Kg in this catchment is to control the volume of water per unit area held in storage and the proportion of direct runoff generated. This will be higher for higher values of Kg. Values for EFF and LF_OBJ are weakly correlated indicating that there is a trade off between the fit at low flows and the overall fit. BEQ95 and EFF are weakly correlated, indicating that there is a tendency to over estimate Q95 when the overall time series fit is optimised.




(a)




(b)

**Figure C.2 Scatter plots for catchment 39098.**

### C.2.3 Sydling Water at Sydling St. Nicholas 44006

The correlation matrix for this catchment is presented in Table C.5 and the scatter plot matrices in Figure C.3a and Figure C.3b. The simple Penman drying curve model predicts that significant soil moisture deficits exist for less that 20% of the time in this catchment. The variability in Cmax is low for this catchment, whereas the variability in b is high. However, b is very small (the 95% value is only 0.16) indicating that the distribution of soil stores within the catchment is strongly skewed towards the value of Cmax. This may

reflect the chalk hydrogeology of the catchment and the associated high infiltration rates and low direct runoff. Alternatively, it may indicate that the model is compensating for errors in input data by maximising the evaporation rate. There is a weaker relationship between Cmax and b than that observed within the previous catchments; an indication of the lesser importance of the loss model in determining mean flow in this higher rainfall catchment. Interestingly Kb is inversely weakly correlated with Cmax, although this may be spurious as the scatter plot indicates that the correlation is limited to not observing high values of Kb for high values of Cmax. Furthermore, the variability in Kg is also relatively small. The loss module parameters, Cmax and Kg, are more identifiable for this catchment than the two dryer catchments. The values of Kg are broadly similar to those observed for the other permeable catchment.

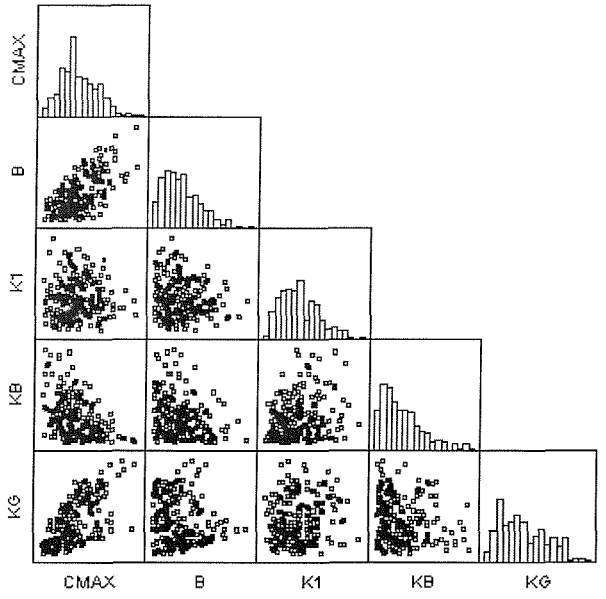
The fit of the model over the lower third of the flow distribution, as represented by LF_OBJ, is dependent on Kg and b. The value of the objective function decreases (improves) with increasing values of Kg, implying a relationship with decreased drainage rates. The correlation with b implies that low values of LF_OBJ only occur for low values of b, whereas high values of LF_OBJ may be obtained for both low and high values of b. BIAS is strongly negatively correlated with Kg. The model will over predict mean flow for lower values of Kg and under predict for higher values.

**Table C.5 Correlation matrix for 44006**

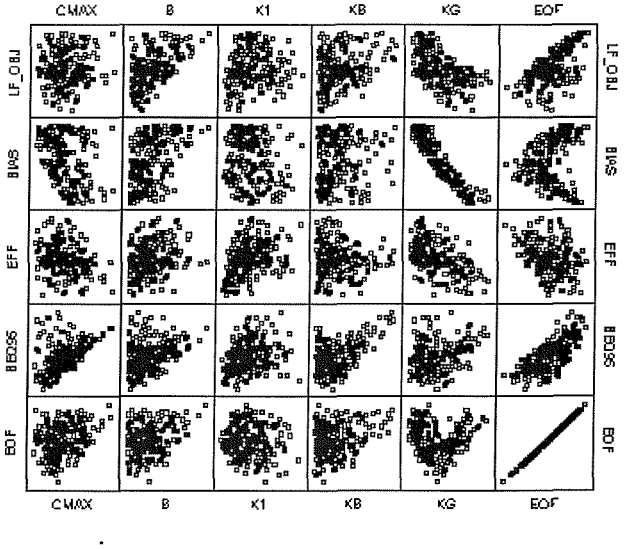
| | CMAX | B | K1 | Kb | Kg | LF_OBJ | BIAS | EFF | BEQ95 |
|---|---|---|---|---|---|---|---|---|---|
| **CMAX** | 1.00 | | | | | | | | |
| **b** | 0.61 | 1.00 | | | | | | | |
| **K1** | | | 1.00 | | | | | | |
| **Kb** | -0.45 | | | 1.00 | | | | | |
| **Kg** | 0.62 | | | | 1.00 | | | | |
| **LF_OBJ** | | 0.50 | | | -0.58 | 1.00 | | | |
| **BIAS** | | 0.45 | | | -0.94 | 0.67 | 1.00 | | |
| **EFF** | | | | | -0.53 | | 0.54 | 1.00 | |
| **BEQ95** | 0.45 | | | 0.49 | | 0.54 | | | 1.00 |

Higher values of Kg would imply reduced drainage, and thus water is likely to be retained in the loss module preventing significant soil moisture deficit from building up resulting in an over-estimation of losses (and hence an under prediction of mean flow). Conversely for low values of Kg evaporation of water may be suppressed due to increased drainage from

the loss module. It should be noted that there is a negative feedback loop as the drainage rate is proportional to the depth of water within the loss module. For a given set of input a higher value of Kg will lead to an enhanced depth of water held in storage which will, in turn, result in a higher drainage rate.




**(a)**




**(b)**

**Figure C.3 Scatter plots for catchment 44006.**

EFF and LF_OBJ are negatively correlated with Kg, again implying a trade off between the overall hydrograph fit (EFF) and the fit at low flows (LF_OBJ), although this is not as

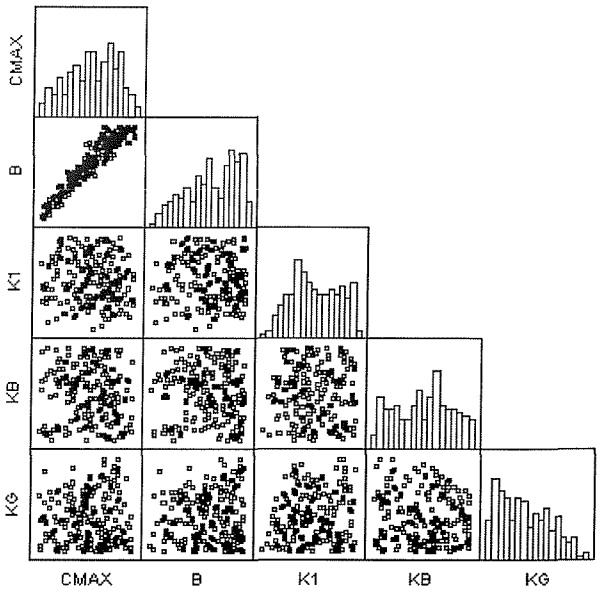
significant as that observed in the other catchments. EFF is positively correlated with BIAS, a function of the correlation of both statistics with Kg. BEQ95, is weakly, and positively correlated to Cmax, Kb and LF_OBJ. The extent of this correlation is that low values of BIAS do not occur at high values of Cmax, Kb or Kg. The fit of the BEQ95 statistic represents a trade off against the LF_OBJ statistic; i.e. a good fit across the lower third of the flow distribution may result in a systematic over estimate of Q95 flows.

### C.2.4 Gelyn at Cynefail 67010

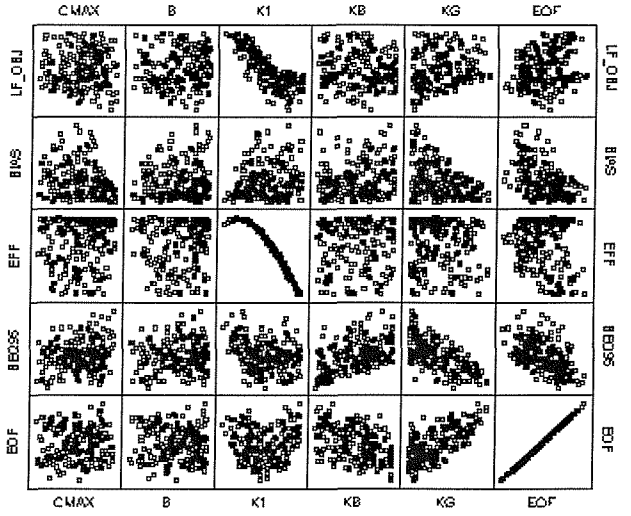
The correlation matrix is presented in Table C.6 for catchment 67010. Scatter plot matrices are presented in Figure C.4a and Figure C.4b. There is generally lower variability in the parameter ranges for 67010 than the other catchments. There is a strong positive correlation between Cmax and b (Table C.6), but no correlations observed between other parameter pairs. The model within this catchment (the BIAS statistic lies between -2.5 and -3%) always underestimates the runoff. The catchment is a very wet catchment; Penman estimated soil moisture deficits occur for about 1% of the time over the calibration period. This underestimate of mean flow is likely to be either a result of an error in the estimated rainfall (too low), the catchment area (too small) or it could be a hydrometric error in the stream flow data. It is unlikely to be an error in the evaporation data as the evaporation demand represents a relatively small fraction of the water balance in this catchment. The model cannot close an exact water balance. The loss model parameters should play a negligible part in the water balance, as significant soil moisture deficits should not build up on a regular basis. However, the low values of Cmax and high values of b observed indicate that soil moisture deficits are likely to build, thus limiting evaporative losses. This may be an example of the model compensating for errors within the input data, or alternatively, this may be a consequence of the thin soils and impermeable substrate within this catchment.

**Table C.6 Correlation matrix for 67010**

| | CMAX | B | K1 | Kb | Kg | LF_OBJ | BIAS | EFF | BEQ95 |
|---|---|---|---|---|---|---|---|---|---|
| **CMAX** | 1.00 | | | | | | | | |
| **B** | 0.95 | 1.00 | | | | | | | |
| **K1** | | | 1.00 | | | | | | |
| **Kb** | | | | 1.00 | | | | | |
| **Kg** | | | | | 1.00 | | | | |
| **LF_OBJ** | | | -0.81 | | | 1.00 | | | |
| **BIAS** | | | | | | | 1.00 | | |
| **EFF** | | | -0.97 | | | 0.75 | | 1.00 | |
| **BEQ95** | | | | 0.48 | -0.65 | | | | 1.00 |




(a)




(b)

**Figure C.4 Scatter plots for catchment 67010.**

The fit at Q95 is good, the within shot variance of BEQ95 is low at 9% with a median value of -7%. This error is comparatively very low, as an absolute error will represent a greater fraction of the observed Q95 in this flashy catchment than for one of the more permeable catchments. The storage capacities within the catchment are low; this causes surface runoff to be generated easily. The values of Kg are very low to compensate for this and enable a realistic base flow contribution to be estimated.

The values of EFF are generally high, although there is a greater variation for this catchment than the other catchments. The value of LF_OBJ is comparable with the other catchments. There is a strong, positive correlation between LF_OBJ and EFF-indicating a trade off again between these objectives functions. This is confirmed by the strong negative correlations between EFF, LF_OBJ and K1, the time constant for the quick flow routing reservoir. High values of K1 give a better fit over the lower third of the flow distribution while low values of K1 are required to model the quick response components of the hydrograph. Kg and Kb are respectively negatively and positively correlated with BEQ95. The errors at Q95 are smaller for a more damped response from the slow flow routing reservoir (as the majority of Q95 bias errors are negative). Smaller errors also occur when the drainage from the loss into the routing reservoir is greater (inversely proportional to Kg) for a given value of S(t).

## C.3 AN EVALUATION OF MODB MODEL FITS WITHIN THE FOUR CASE STUDY CATCHMENTS

The number of catchments for which an acceptable water balance could be closed was reduced from 179 for MODA to 171 for MODB. One of the catchment for which a water balance could not be closed was 44006. The 95$^{th}$, median and 5$^{th}$ percentile parameter and objective function limits for MODB are presented in Table C.7 for the four case study catchments. The significant correlations of greater than 0.4 are presented in Table C.8. The parameter ranges for K1 and Kb are broadly similar to those observed for MODA. The range is much less for Cmax (Cmax is much more identifiable) and the median values for Cmax are much smaller than those observed for MODA, a consequence of not having the soil drainage component within the model and the assumption of a uniform distribution. The loss module configuration for MODB is analogous to a probability distributed form of

the Penman drying curve model. The values of LF_OBJ and EFF (with the exception of 67010) and BEQ95 are all substantially worse demonstrating that poorer fits are obtained with MODB, particularly for EFF in 34003. The values of EFF obtained with MODB for 67010 are substantially higher than those obtained with MODA.

**Table C.7 Ranges for parameters and objective function values for the test catchments using MODB**

| | **34003** | | | **39098** | | | **67010** | | |
|---|---|---|---|---|---|---|---|---|---|
| % | 95 | 50 | 5 | 95 | 50 | 5 | 95 | 50 | 5 |
| Cmax | 257 | 237 | 217 | 422 | 381 | 354 | 329 | 129 | 39 |
| K1 | 242 | 107 | 34 | 38 | 24 | 12 | 26 | 18 | 6 |
| Kb | 1783 | 1515 | 1147 | 402 | 285 | 114 | 491 | 446 | 404 |
| LF_OBJ | 44 | 39 | 35 | 88 | 76 | 72 | 70 | 59 | 50 |
| BIAS | 3 | 1 | -1 | 3 | 0 | -3 | 2 | -1 | -3 |
| EFF | 0.15 | .16 | 0.0 | 0.74 | 0.71 | 0.53 | 0.81 | 0.79 | 0.69 |
| BEQ95 | -45 | -53 | -65 | -20 | -39 | -60 | 34 | 13 | -8 |

The parameters are uncorrelated with one another in all catchments. BIAS is strongly correlated with Cmax, as would be expected. In the permeable catchment (34003) EFF, LF_OBJ and BEQ95 are strongly correlated with Kb, and the value of both objective functions improves with increasing Kb. EFF is strongly correlated with K1 in the impermeable catchments (39098 & 67010). For 39098, the LF_OBJ function also correlates with K1and BEQ95 is also strongly correlated with Kb. There is no evidence of the trade off between EFF and LF_OBJ observed with MODA within these catchments. There is therefore no evidence of trade off between the overall fit of the hydrograph and the fit at low flows for MODB. This is a consequence of the use of a fixed split for partitioning effective rainfall.

**Table C.8 Parameter and Objective Function correlation matrices for MODB within the case study catchments**

| | Cmax | K1 | Kb | LF_OBJ | BIAS | EFF | BEQ95 |
|---|---|---|---|---|---|---|---|
| **34003** | | | | | | | |
| Cmax | 1.00 | | | | | | |
| K1 | | 1.00 | | | | | |
| Kb | | | 1.00 | | | | |
| LF_OBJ | | | -1.00 | 1.00 | | | |
| BIAS | -1.00 | | | | 1.00 | | |
| EFF | | | 0.77 | -0.78 | | 1.00 | |
| BEQ95 | | | 0.98 | -0.98 | | 0.64 | 1.00 |

| | Cmax | K1 | Kb | LF_OBJ | BIAS | EFF | BEQ95 |
|---|---|---|---|---|---|---|---|
| **39098** | | | | | | | |
| Cmax | 1.00 | | | | | | |
| K1 | | 1.00 | | | | | |
| Kb | | | 1.00 | | | | |
| LF_OBJ | | 0.46 | -0.86 | 1.00 | | | |
| BIAS | -1.00 | | | | 1.00 | | |
| EFF | | 0.86 | | | | 1.00 | |
| BEQ95 | | 0.49 | 0.76 | -0.46 | | | 1.00 |

| | Cmax | K1 | Kb | LF_OBJ | BIAS | EFF | BEQ95 |
|---|---|---|---|---|---|---|---|
| **67010** | | | | | | | |
| Cmax | 1.00 | | | | | | |
| K1 | | 1.00 | | | | | |
| Kb | | | 1.00 | | | | |
| LF_OBJ | | | | 1.00 | | | |
| BIAS | -0.79 | | | | 1.00 | | |
| EFF | | 0.76 | | -0.81 | | 1.00 | |
| BEQ95 | 0.74 | | | 0.64 | -0.61 | | 1.00 |

# Appendix D:
# A Region of Influence approach for regionalising MODA

# D A Region of Influence approach for regionalising MODA

## D.1 INTRODUCTION

The quality of model fits, obtained when MODA and MODB were run with model parameters predicted using the regression models was extensively evaluated. This evaluation, presented in detail within Chapter 8, demonstrated that the fits obtained with the regression based regionalised model parameters were generally of a higher quality for the simpler model configuration MODB. However, the model structure is not as conceptually attractive as that for MODA. This is due to the more simplistic treatment of soil moisture behaviour and the partitioning of effective rainfall between quick and slow flow routing paths. Considering the performance of the calibrated model parameters, MODA was found to be more effective in closing a water balance in a greater number of catchments, although there is strong evidence of the flexibility of the model structure compensating for errors in climatic data.

One potential reason for the performance of the regionalised parameters for MODB being greater than that of their MODA counterparts is the strong co-linearity observed between the MODA loss module parameters. This co-linearity is ignored in the regression based regionalisation analysis, in which each parameter is treated as being independent of the others within the vector. An alternative approach to regionalisation was developed based on retaining the parameter vector as a single entity, and then using suitable catchment characteristics to select the most appropriate calibrated parameter vectors from the pool of catchments for application at an ungauged site. The type of analysis, called a Region of Influence (ROI) has been used in catchment classification (Burn and Boorman, 1993) and has been used by a number of researchers within regional flood frequency estimation (e.g. Burn and Goel, 2000; Robson and Reed, 1999). This approach was not applied to MODB, as there was negligible correlation between the parameters for this model configuration. The selection of catchment characteristics for assessing catchment similarity is presented withinD.1.1 . The ROI algorithm developed for the study is presented within D.1.2. This section also includes a comparison of the performance of the ROI approach in predicting

stream flow with that of the regression based approach for MODA.

### D.1.1 Selection of characteristics for assessing similarity

Canonical Correlation Analysis (CCA), Manly (1994), was used to select the most appropriate catchment characteristics for indexing the entire model parameter vector. CCA is used to identify and optimise the correlation between linear combinations (V) of dependent (Y) variables and linear combinations (U) of independent (X) variables. It can be thought of as an extension of multiple regression in which several Y variables are simultaneously related to several X variables. For $p$ X variables and $q$ Y variables there can be up to $r$ pairs of linear combinations (where $r$ is equal to the smaller of $p$ or $q$) that are significantly correlated. These $r$ pairs are chosen so that the correlation between U1 and V1 is a maximum, U2 and V2 is a maximum, subject to being un-correlated with U1 and V1, and so forth. The first pair will have the highest correlation, the second the second largest correlation, etc.

Solving the following Eigen value problem identifies canonical variables and associated correlations:

$$\left(B^{-1}C'A^{-1}C-\lambda I\right)b=0, \qquad \textbf{(D.1)}$$

where:
B= $q$x$q$ dependent variable correlation matrix;
A= $p$x$p$ independent variable correlation matrix;
C= $p$x$q$ correlation matrix between dependent and independent variables;
I = an identity matrix.

X and Y are standardised to have zero mean and unit variance. The eigen values are the squares of the correlation between the canonical variable pairs. Th eigen vectors b1,b2,...br give the coefficients of the Y variables for the canonical variables. The coefficients of Ui, the ith canonical variable for the X variables, are given by the elements of the vector:

$$a_i = A^{-1}Cb_i. \qquad \textbf{(D.2)}$$

As with PCA, the interpretation of coefficients for X and Y for a particular canonical variable pair is problematical because of the potential covariance problems between elements of the X vector, or the Y vector. For this reason loadings, which are the correlation between the canonical variable in question and each of the component variables, are often used. The plot axes for the canonical variables can be rotated if a variable has a sizeable loading for more than one canonical variable. The aim is to maximise the component loading for a variable with one PC and minimise it with respect to the others.

The catchment characteristics used in the analysis were those continuous variables that were found to be significant in determining the regression based relationships (Chapter 7). Variable transformations that were found to be useful in the regression analysis were applied to both the catchment characteristics and the calibrated model parameters. The Y variables considered were therefore log($\overline{C}$ +1), b, log(Kg+1), log(K1+1) and Kb. The X variables considered were log(PP+1), log(LDP+1), log(AREA+1), log(DPL_BAR+1), log(HOSTRES+1) and HOSTBFI. Initial analysis identified that there was high redundancy in the use of log (LDP+1), log(AREA+1) and log(DPL_BAR+1) due to the high inter-correlation between these variables. For this reason log(LDP+1) and Log(DPL_BAR+1) were removed from the analysis, as the correlation between Log(AREA+1) and the model parameters tended to be greater.

The resultant analysis is summarised in Table D.1. The canonical variables have been rotated to aid the interpretation of the loadings. The total shrunk (or population estimate) $R^2$ of 0.88 indicates that the set of three independent canonical variables explains 88% of the variation in the set of dependent canonical variables. The first entries in the table list the correlation between the canonical variable pairs, both before and after rotation. This demonstrates that the first two pairs dominate the relationships. Looking at the coefficients and rotated loadings for these pairs in more detail, it can be seen that Kb and log (K1+1) dominate the dependent CV in the first pair, and the independent CV is dominated by BFIHOST. This confirms the regression relationships for these parameters, that these variables are dominated by the hydrogeological nature of the catchment, as represented by BFIHOST.

**Table D.1 Summary of the canonical correlation analysis**

| | | | Canonical Variables | | |
|---|---|---|---|---|---|
| | | | 1 | 2 | 3 |
| | **Canonical:** | **Unrotated** | 0.83 | 0.74 | 0.40 |
| | **Correlations:** | **Rotated** | 0.74 | 0.64 | 0.30 |

| **Dependent** | **variables:** | **1** | **2** | **3** | **Independent Variables:** | | **1** | **2** | **3** |
|---|---|---|---|---|---|---|---|---|---|
| *Coefficients* | B | -0.09 | -0.03 | 0.40 | *Coefficients* | BFIHOST | 0.96 | 0.48 | -0.37 |
| | KB | 0.24 | 0.14 | 0.91 | | LOG_PP | 0.08 | -1.06 | 0.35 |
| | LOG_K1 | 0.43 | -0.06 | 0.38 | | LOG_AREA | 0.05 | 0.18 | 1.01 |
| | LOG_KG | -0.52 | 1.35 | 0.90 | | | | | |
| | LOG_CBAR | 0.83 | -0.57 | -1.27 | | | | | |
| *Rotated* | B | -0.15 | **-0.74** | 0.29 | *Rotated* | BFIHOST | **0.97** | -0.22 | 0.09 |
| *Loadings* | KB | **0.86** | -0.02 | -0.03 | *Loadings* | LOG_PP | 0.22 | **-0.97** | -0.04 |
| | LOG_K1 | **0.72** | 0.24 | -0.22 | | LOG_AREA | 0.08 | 0.04 | **0.99** |
| | LOG_KG | 0.03 | **0.99** | -0.11 | | | | | |
| | LOG_CBAR | 0.29 | **0.64** | **-0.66** | | | | | |
| | Total Shrunk $R^2$ | | 0.88 | | | | | | |

In the second pair, the dependent CV is dominated by the loss model parameters whilst the independent CV is dominated by log (PP+1), the logarithm of the potential for the catchment to generate significant soil moisture deficits. The negative loading for the Pareto shape parameter, b, and the positive loadings for log ($\overline{C}$ +1) are understandable. For a given value of Cmax, a lower value of b will give rise to a higher value of $\overline{C}$. This result indicates that the role of parameters in controlling evaporation is more important within high PP catchments. The positive loading for Kg supports this theory. The nett effect of a higher value of Kg will be to retain water within the loss module leading, potentially, to an enhanced evaporation loss. The third pair of CV are not highly correlated. The pair is dominated by Log ($\overline{C}$ +1) within the dependent variable and Log (AREA+1) within the independent variable. This indicates a weak relationship between $\overline{C}$ and catchment size, as was found to be the case in the regression modelling of Cmax for MODB. On the basis of this canonical correlation analysis, BFIHOST and PP were selected as the catchment characteristic indices that dominate the relationships between the variation in parameter vectors and the variation within the continuous catchment characteristics.

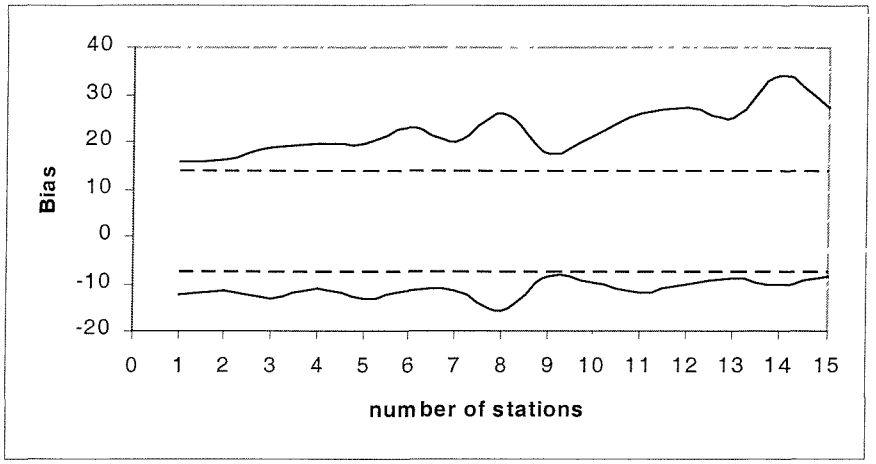
### D.1.2 The ROI algorithm

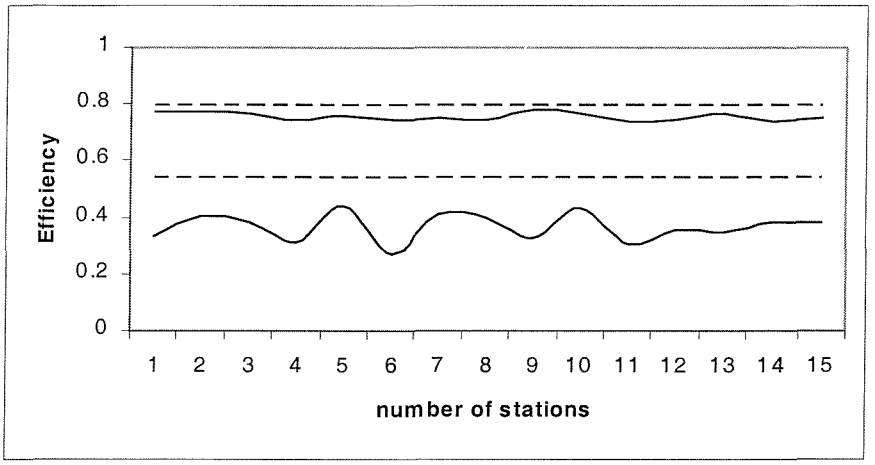
The utility of three schemes was evaluated for selecting analogue catchments, based on the similarity in BFIHOST and PP from a source pool of catchments. These schemes were evaluated by considering each catchment from the data set of *m* calibrated catchments for MODA in turn -termed the target catchment - and then selecting *n* nearest neighbours from the remaining pool of *(m-1)* catchments- termed the source catchments. The model for the target catchment was run using calibrated parameters from each of the n catchments and the catchment area and climatic data for the target catchment. This process yields *n* simulated time series of flows for the target catchments. Taking an arithmetic average to give the final simulated flow time series for the un-gauged catchment combined these time series. The fit of this time series was then evaluated against the gauged flow time series. The three schemes for selecting nearest neighbour catchments were:

1. To rank the m-1 source catchments in order of absolute difference in HOSTBFI between the source gauge and the target gauge. To then select the gauges from the pool of (m-1) for which the difference in HOSTBFI was less than 0.07 (The 68% prediction limit for the HOSTBFI regression model). This identified K catchments, which do not have significantly different BFIHOST values at the 68% confidence level. These K gauged were then ranked in order of the absolute difference in PP. Taking the n nearest neighbour catchments in terms of PP from these K source catchments then identified the ROI for the target gauge.
2. As for (a) but using the 95% prediction limit for the HOSTBFI model assuming normality.
3. As for (a) but using a further condition that the difference in catchment area should not exceed 20%. The sensitivity of the results to the value of this condition was evaluated.

These three schemes gave very similar results. The 68% confidence interval is plotted as a function of n for the goodness of fit statistics BIAS, EFF and BEQ95 for scheme (a) in Figure D.1 over the pool of m catchment during the calibration period. The 68% confidence intervals obtained for MODA using the regression based model parameter estimates are also included within the plots as dashed lines.

This figure demonstrates that the BIAS 68% confidence interval for the ROI approach is larger than that for the regression based parameter estimates, and it increases in width as n increases. The EFF confidence interval is both wider than the regression based estimates and corresponds to lower median values of EFF. For some values of n, the BEQ95 confidence interval is similar to that for the regression based estimates. In general though, the confidence interval is wider. The differences in BIAS, $R^2$, EFF and BEQ95 between the results obtained using the ROI approach for n=1 and the regression based model parameter estimates are summarised as histograms in Figure D.2. Within the figure negative values indicate degradation in performance when using the ROI based estimates. These histograms confirm that the ROI based approach is marginally worse for BIAS, significantly worse for EFF and similar in performance for BEQ95. Given these results the ROI approach was not considered further.



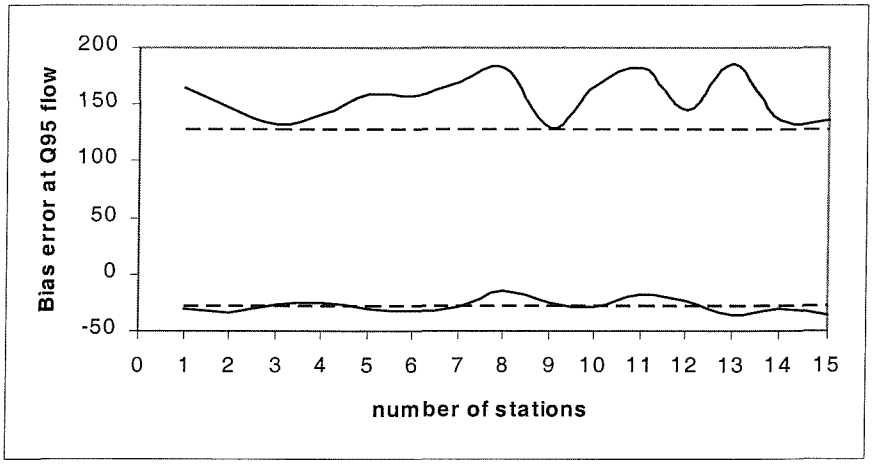




**Figure D.1 Comparison of the performance ROI and regression based parameter estimates.**

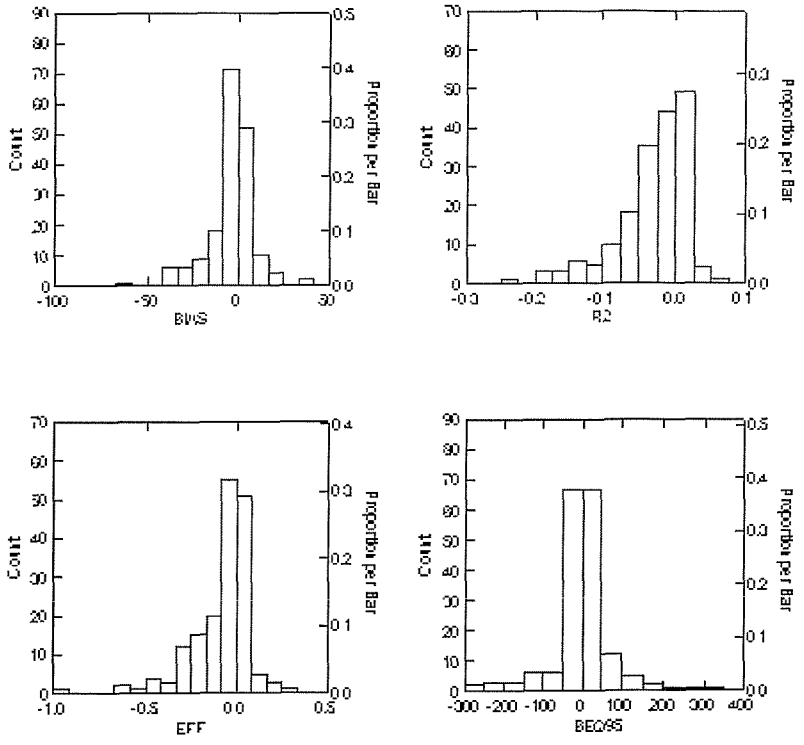



**Figure D.2 Comparison of ROI results and regression based results for MODA.**

# Appendix E:
# The development of regression models for regionalising MODA and MODB: Results

# E The development of regression models for regionalising MODA and MODB: Results

The full results for the development of regression models for predicting the rainfall- runoff model parameters are presented within this appendix. The results are summarised within Chapter 7. The results of the regression analysis are presented by parameter for each model. In each case, a summary table describing the structure of the 100% model is presented. This is accompanied by plots of residuals and the predicted values as a function of the observed values. In each case outliers are identified, and the quality of the model fit and physical interpretation of the model structure are discussed.

## E.1 REGRESSION MODELS FOR PREDICTING MODA PARAMETERS

### E.1.1 The maximum storage capacity: Cmax

Correlation analysis coupled with un-transformed and transformed variable regression analysis demonstrated that there were no significant relationships between Cmax and any of the candidate catchment characteristics. To investigate whether this was a consequence of the strong covariance between Cmax and b (Pareto shape parameter) in many of the catchments, Cmax was expressed as the mean storage capacity, $\overline{C}$, derived by expressing Cmax as a fraction of (b+1) as discussed in Chapter 4. The resultant distribution of $\overline{C}$ values was positively skewed. This is in part due to large Cmax and low b values in some catchments, which may be compensating for errors within the climatic data inputs as discussed in chapters 2 (in the context of the literature on this subject) and 6 (in the context of the case study catchments). To address this skew, the distribution was normalised by taking logarithms of ($\overline{C}$ +1). The histograms of un-transformed and transformed $\overline{C}$ values are presented in Figure E.1.

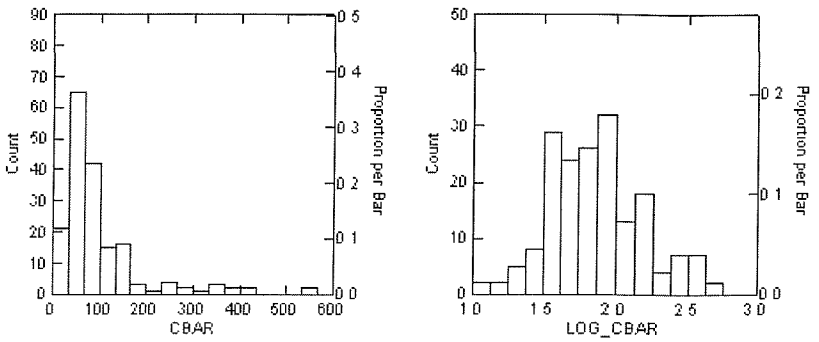



**Figure E.1 Histograms for $\overline{C}$ and log($\overline{C}$ +1).**

**Table E.1 Regression model for log ($\overline{C}$ +1)**

| Cbar_HG | HOST Classes | Hydrogeological units | Coefficient | Std error | Tolerance | P |
|---|---|---|---|---|---|---|
| 1 | 1 | Chalk, chalk drift | 0.026 | 0.001 | 0.932 | 0 |
| 2 | 2 | Oolitic limestone, soft magnesian | 0.019 | 0.001 | 0.913 | 0 |
| 3 | 5, 7 | Blown sand, gravels | 0.023 | 0.002 | 0.758 | 0 |
| 4 | 6 | Colluvium, coverloam, gravel, loamy drift | 0.019 | 0.002 | 0.849 | 0 |
| 5 | 20, 23, 25 | Very soft massive clays | 0.02 | 0.001 | 0.816 | 0 |
| 6 | 17, 19, 22, 27 | Hard rock | 0.019 | 0.001 | 0.796 | 0 |
| 7 | 16, 18, 21 | Soft bedded clays, loams, till, shales | 0.02 | 0.001 | 0.861 | 0 |
| 8 | 24, 26 | Soft shales, siltstones, bedded clays/loams, clay and flints | 0.017 | 0.001 | 0.722 | 0 |
| 9 | 12, 28, 29 | Weathered intr/meta. rock, raw peats, eroded peats | 0.008 | 0.001 | 0.721 | 0 |
| 10 | 3 | Soft sandstone | 0.027 | 0.002 | 0.858 | 0 |
| 11 | 4 | Weathered intr/meta. rock, hard limestone and sandstone | 0.023 | 0.002 | 0.924 | 0 |
| 12 | 14,15 | Weathered intr/meta. rock, colluvium, coverloam, loamy,drift gravels/loams, sandstone | 0.016 | 0.001 | 0.775 | 0 |
| 13 | 13,8, 9, 10,11 | Weathered intr/meta. rock, chalky drift, loams Earthy peats, Shattered rock, alluviums, cover loam, chalky drift | 0.012 | 0.005 | 0.685 | 0.01 |
| | | adjusted $R^2$= 0.988 | | SE= 0.199 | | N=177 |

The best regression relationship between log ($\overline{C}$ +1) and catchment characteristics was based on a grouping of HOST called Cbar_HG. Two potential outliers were identified catchments 54022 the Severn at the Plynlimon Flume and 84022. Duneaton at Maidencots. These catchments are dominated by HOST class 15, which consists of weathered rock,

sandstone, colluvium and gravels/loams. On inspection, other catchments dominated by HOST 15 tend to have high values of b indicating the distribution soils is skewed toward shallow storage capacities, as would be expected. The two outliers identified have very low b values of 0.013 and 0.176 (respectively) and consequently high values of $\overline{C}$. From the analysis of model behaviour presented in Chapter 6 this is normally associated with errors in climatic data/and or catchment area (over estimation of catchment rainfall) and the response of the model in trying to close a water balance by maximising the evaporation of water at the potential rate. The Plynlimmon experimental catchment is very small and lies in an area of heterogenous rainfall fields. The Duneaton catchment, on inpection, lies in an area with a very sparse rain gauge network. These catchments were therefore removed from the model.

The plot of residuals, Figure E.2, and of predicted versus observed log ($\overline{C}$ +1), Figure E.3a, demonstrate that the model tends to under predict for high values of $\overline{C}$. Furthemore, there are some large negative (under estimates) around observed log($\overline{C}$ +1) values of 2.0-2.5 ($\overline{C}$ values between 100 and 300mm). Inspection of these catchments identified that these were all high rainfall catchments in which the optimal value of b is low compared within similar catchments, and the model is thus likely to be compensating for errors in the climatic data.

The groupings, regression coefficients and model fit statistics for all catchments are presented in Table E.1. The grouping of HOST classes is based on the soil integrated air capacity (as a surrogate for permeability) and geological substrate. Permeable soils, particularly when underlain by permeable geologies have higher coefficients and hence higher values of $\overline{C}$. This is conceptually correct as these soils will have high infiltration rates and will have a greater storage capacity. Conversely, thin impermeable soils or thin soils underlain by impermeable geologies will have low storage capacities. The model is essentially a weighted average of percentage coverages of the HOST groupings with the weight given by the coefficient estimate. This is conceptually attractive, as $\overline{C}$ is an average value of storage capacities. The logarithm of ($\overline{C}$ +1) is used within the regression so that the model can predict a zero value of $\overline{C}$ for a catchment with zero fractional extents. This cannot occur in practice, as it implies a zero catchment area, but it is mathematically correct.

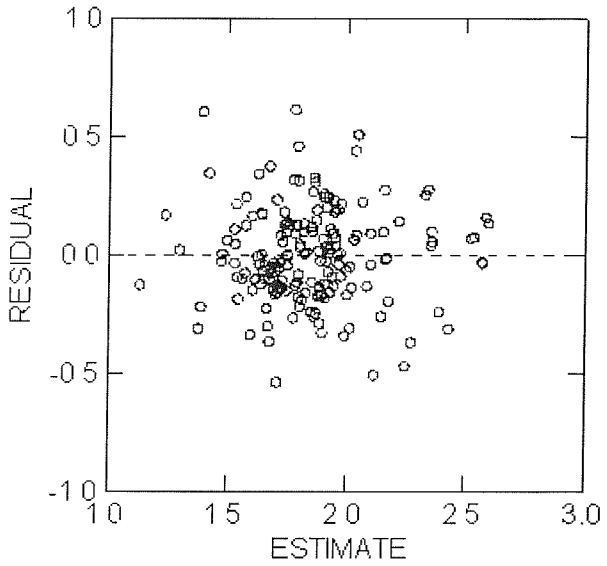
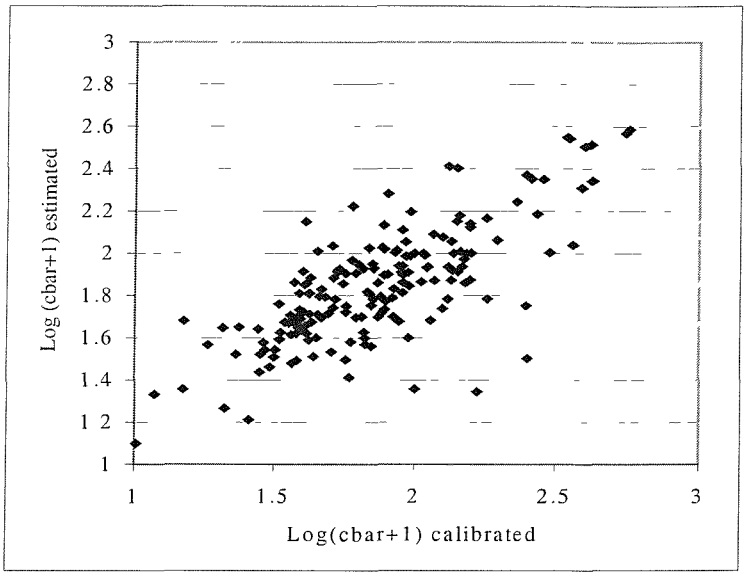


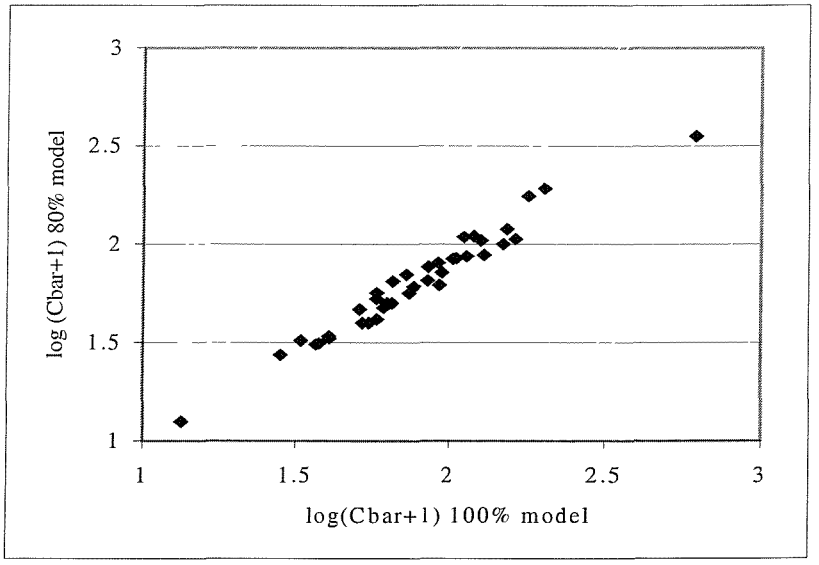

**Figure E.2 Regression residuals for log ($\overline{C}$ +1).**

All parameter coefficients are significant at the 95% confidence level and the standard errors are low, although this in part a consequence of the model being constrained to pass through the origin. The tolerance values are high indicating that there are no covariance/ redundancy issues within the model. The adjusted $R^2$ value is very high, a function of the fact the model has a zero intercept and therefore is not constrained to pass through the mean of the data. Of more importance is the Standard Error (SE) of 0.199. This equates to a Factorial Standard Error of 1.58, that is (given N is reasonably large) the predicitve accuracy of the model is approximately ±58% at the 68% level.

The comparison of estimates generated using the same model calibration on 80% of the catchments for the remaining 20% (42) catchments (Figure E.3b) demonstrate that the model is stable. This graph, which is for one of the 80:20 split tests is included as an example of this type of plot. The observed scatter is representative of that observed for regression models based on the discrete HOST classes, irrespective of the rainfall runoff model parameter being modelled. The small scatter observed with the HOST based relationships is due to the low representation of particular HOST classes. The sensitivity of the model coefficients for these classes is sensitive to the fractional representation of these classes within the 80:20 split samples.




(a)




(b)

**Figure E.3 Graphs of model fit for log ($\overline{C}$ +1).**

### E.1.2 The Pareto shape parameter, b

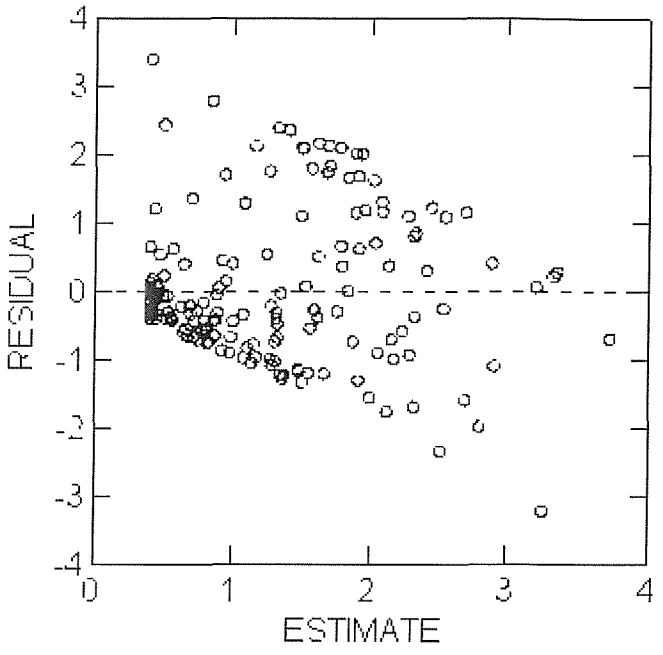
The optimal model for the Pareto shape parameter b was a linear model that related b to the fractional extent of a grouping of HOST classes, b_HG. It was not possible to obtain a significant relationship between b and the continuous characteristics that explained more than 10% of the variance. The grouping, regression coefficients and model fit statistics for all catchments are presented in Table E.2. The 80:20 split sample tests demonstrated that the model is stable.

**Table E.2 Model for the Pareto shape parameter, b**

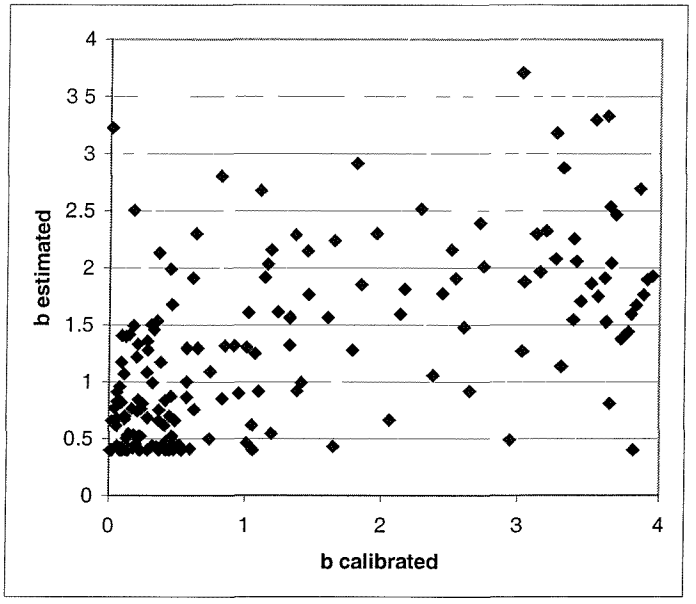
| b_HG | HOST Classes | Hydrogeological units | Coefficient | Tolerance | P |
|---|---|---|---|---|---|
| 1 | 15 | Weathered intr/meta Rock, sandstone, colluvium, gravels/loams | 0.023 | 0.781 | 0 |
| 2 | 4, 5, 7,6,24, 26 | Weathered intr/meta Rock, hard limestone and sandstone Blown sand, gravels Colluvium, coverloam, gravel, loamy drift Soft shales, siltstones, bedded clays/loams, clay and flints Weathered int/meta Rock, colluvium, coverloam, loamy drift | 0.007 | 0.769 | 0 |
| 3 | 1,2,13,8,9, 10, 20, 23, 25,16, 18, 21, 3 | Chalk, chalk drift Oolitic limestone, soft magnesian Weathered mtr/meta rock, chalky drift, loams Shattered rock, alluviums, cover loam, chalky drift Very soft massive clays Soft bedded clays, loams, till, shales Soft sandstone | 0.004 | 0.956 | 0.01 |
| 4 | 12, 28, 29,11 | Weathered intr/meta Rock, raw peats, eroded peats Earthy peats | 0.052 | 0.715 | 0 |
| 5 | 17, 19, 22, 27 | Hard rock | 0.014 | 0.867 | 0 |
| adjusted $R^2$= 0.658 | | SE= 1.046 | | | N= 179 |

The regression model fit is very poor, probably due in part to the covariance between Cmax and b. The residuals plot, Figure E.4a and plot of predicted against observed, Figure E.4b, indicate that the model fails to adequately describe the variance within the data set. The standard error is 1, giving an approximate predictive accuracy of ±1 at the 68% level. The mean and standard deviation of the raw observed data set are 1.2 and 1.3 respectively. This indicates that the model does not give a major improvement over just using the mean, although it should be noted that the parameters are all significant. Outliers were not identified, due to the lack of variance explained by the model.

The optimal grouping of HOST is a 6-class grouping. The grouping was obtained using the grouping strategy used for the mean storage capacity $\overline{C}$. Thin soils overlying hard rock substrate have a high value of b (implies a distribution skewed towards the shallow soil storage capacities). The model predicts that the distribution of soil capacities on permeable soils, particularly when overlying permeable substrates, will be skewed towards the maximum soil capacity (low value of b). The implications of this are that the impermeable soils will start to generate direct runoff earlier than permeable soils for a given precipitation event. Furthermore significant soil moisture deficit will occur more frequently in these lower storage soils. This is conceptually correct and is consistent with the observations of model behaviour of the model within the case study catchment analysis presented within Chapter 6.




**(a)**




**(b)**

**Figure E.4 Graphs of model fit for the Pareto shape parameter, b.**

### E.1.3 The loss model drainage constant, Kg

The optimal model identified for the drainage constant was one relating the logarithm of (Kg+1) to the logarithms of PP, LDP and HOSTRES. A logarithmic model was used as Kg is skewed and a logarithmic transformation was found to optimise the correlation between Kg and the dependent variables. One was added to each of the variables prior to taking the logarithms, as a zero value is feasible for all of the elements of the model; the logarithm of which is infinity. Two outliers were removed from the model: 31023, West Glen at Easton Wood and 39017, the Ray at Grendon Underwood. These catchments are both small (4.4 and 18.6 $km^2$ respectively), clay (respectively Boulder and Oxford) catchments. The daily flow records showed that both catchments were ephemeral, drying up in the summer months. The model is not structured to model an ephemeral stream as it assumes a non-zero outflow from the routing reservoirs and thus these catchments were removed from the analysis.

**Table E.3 Regression model for log (Kg+1)**

| Variable | Coefficient | Std error | Tolerance | P |
|---|---|---|---|---|
| CONSTANT | 3.341 | 0.114 | . | 0 |
| LOG_PP | 8.593 | 0.663 | 0.979 | 0 |
| LOG_LDP | -0.306 | 0.074 | 0.99 | 0 |
| LOG_HOSTRES | -4 .257 | 1.091 | 0.989 | 0 |
| Adjusted $R^2$= 0.58 | | SE= 0.23 | | N=177 |

The regression coefficients and model fit statistics for the all catchments model are presented in Table E.3. All coefficients are significant at the 95% level. The independent variables are all orthogonal (tolerance values of >0.9). The 80:100 split model tests demonstrated that the model was very stable. The two models were very stable with a near one to one relationship over the 42 catchments within the 20% samples. A example of the relationships between the values of log (Kg+1) predicted for an independent 20% test using the 100% model and an 80% model is presented within Figure E.5. This is a typical example of this relationship for models based upon continuous variables, irrespective of the rainfall runoff model parameter being estimated.

The plot of residuals and the fit of the model are presented in Figure E.6a and b respectively. These indicate that the model has a tendency to over-predict at low values and under-predict at high values. The FSE for the model is 1.70 and the model explains 58% of the variance.

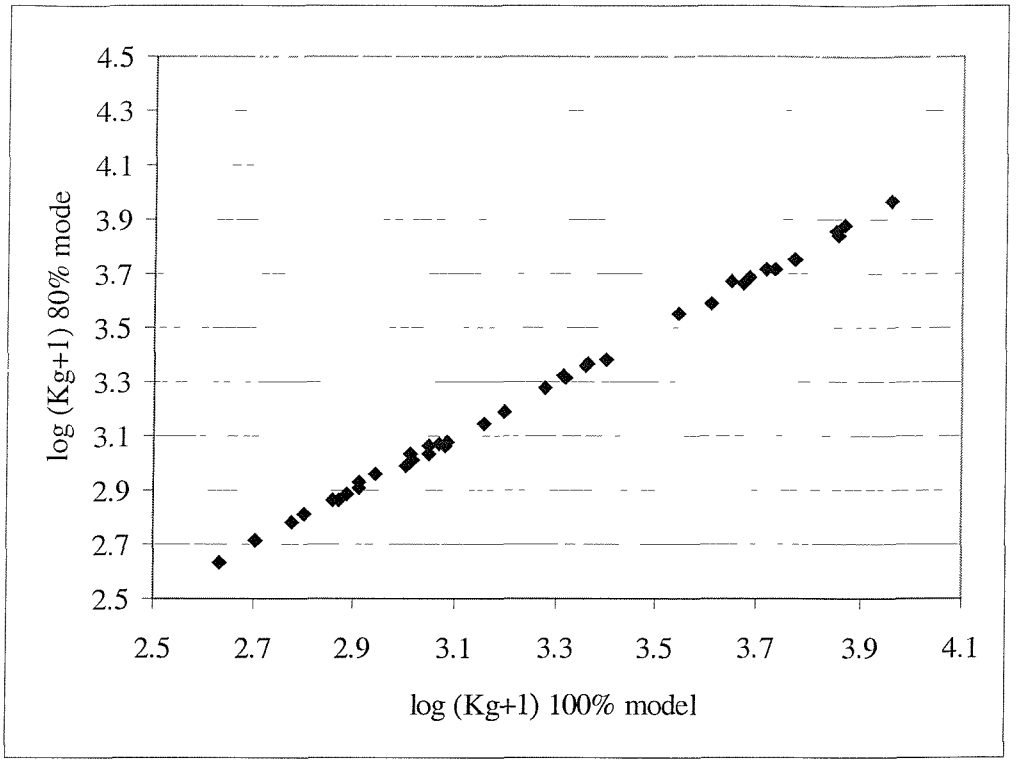
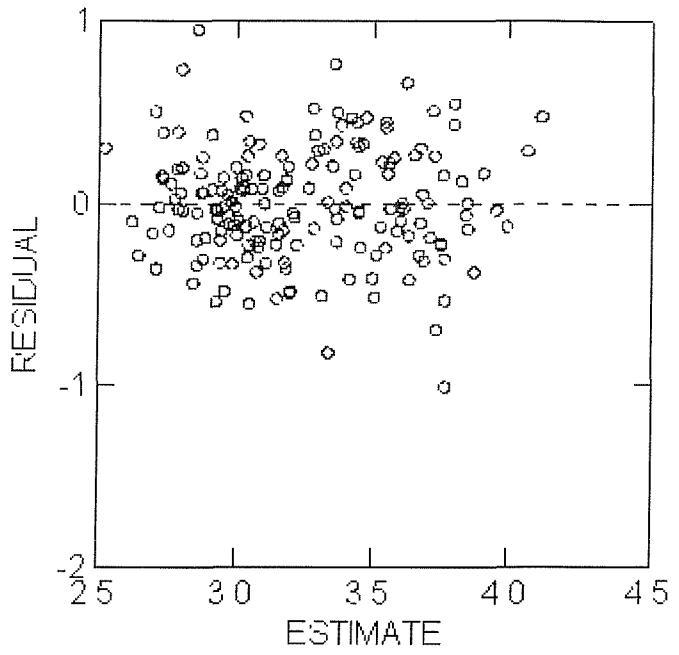
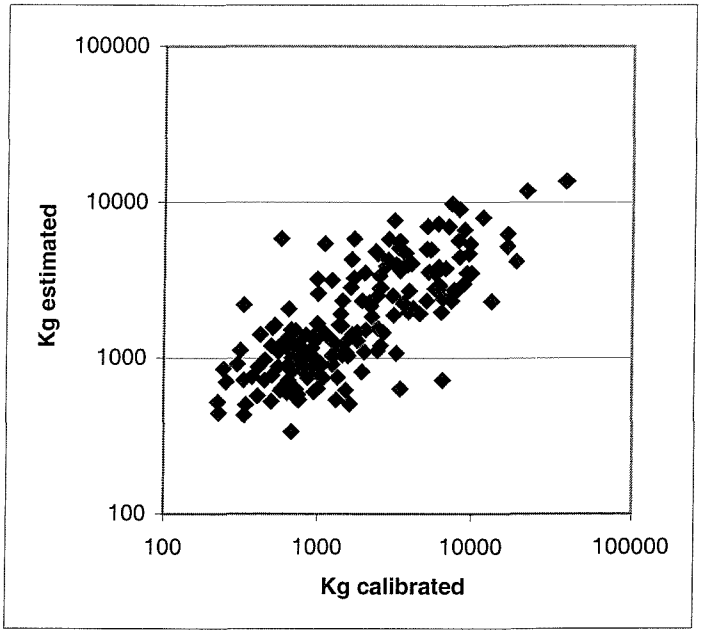



**Figure E.5 Graph of the stability of the log (kg+1) model.**

The model implies that Kg will be larger and therefore limiting drainage, for catchments that have a tendency to build up soil moisture deficits, as represented by PP. The negative coefficient for LDP indicates Kg is smaller in large catchments, which implies that drainage, and hence base flow, is potentially greater in larger catchments. This is an indication that the model is representing the averaging effects of larger catchments by increasing the drainage from the soil store, and hence the fraction of effective rainfall routed through the slow flow routing reservoir. Kg decreases with increasing HOSTRES. A high value of HOSTRES indicates that the catchment has a higher BFI (and hence base flow) than would be expected based upon anticipated SPR. This again is intuitively correct.




**(a) Residuals**




**(b) Predicted values of kg as function of calibrated values**

**Figure E.6 Graphs of model fit for the drainage constant, Kg.**

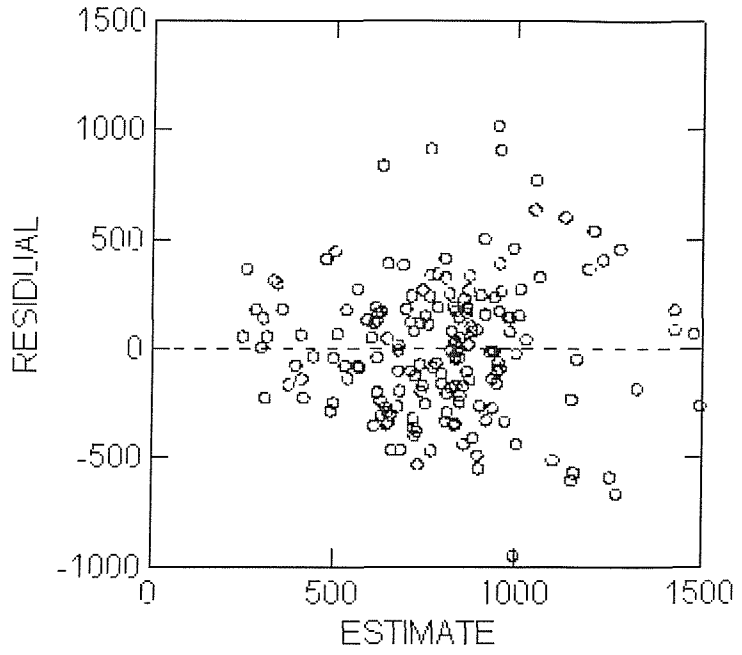
### E.1.4 The slow flow routing reservoir time constant, Kb

The optimal, stable model identified for this parameter was one that relates Kb to BFIHOST and the logarithm of AREA. The regression coefficients and model fit statistics for all catchments are presented in Table E.3. All coefficients are significant at the 95% level. The independent variables are all orthogonal (tolerance values of >0.9).

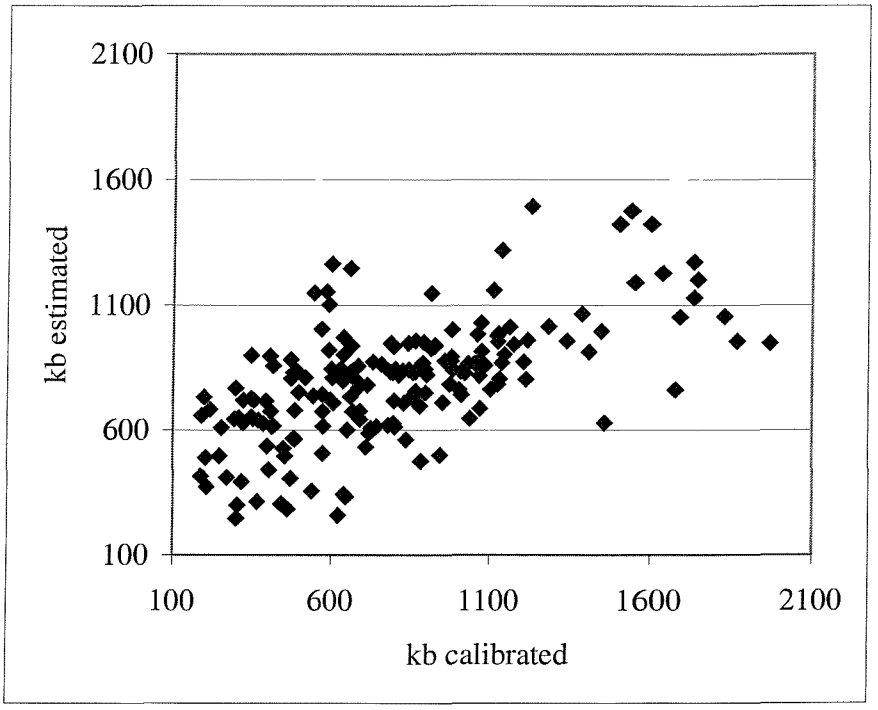
**Table E.4 Regression model for Kb**

| Variable | Coefficient | Std error | Tolerance | P |
|---|---|---|---|---|
| CONSTANT | -260.0 | 110.3 | . | 0.019 |
| LOG_AREA | 170.1 | 40.9 | 0.965 | 0 |
| BFIHOST | 1389.2 | 167.6 | 0.965 | 0 |
| | adjusted $R^2$= 0.51 | | S.E.=298 | N=179 |

LOG_AREA was used as area (as with LDP) has a very skewed distribution. The average annual rainfall SAAR was found to be significant when included within the model. However, it was omitted from the final model as, although the coefficient was significantly different from zero at the 95% limit, the condition indices between it and the constant were very high and furthermore the parameter estimate was low. Removal of SAAR reduced the $R^2$ and for the model from 0.465 to 0.46 and increased the SE by 2. The model has a relatively low $R^2$ and from the plots of residuals (Figure E.7a) and the plot of estimated as a function of calibrated values (Figure E.7b) it can be seen that the model significantly under predicts for high observed values of Kb. An alternative model was identified using the fractional extents of HOST classes (using a grouping based on hydrogeological units for the soil classes). The fit of this model was similar at high values of Kb to the one presented but grossly over estimated at low values of Kb. The final model predicts higher values of Kb for high base flow (BFI) and larger catchments. This is conceptually correct. Permeable systems are high storage systems in which stream flow is dominated by release of water from groundwater with a generally small direct runoff component. As a result rates of recession rates are low in these catchments, reflected in the model by higher values of Kb. In larger catchments both hill slope and groundwater routing will become more damped and channel routing and storage effects may start to become important. One aspect of this is to reduce the variance of stream flow data and rates of changes within stream flow. A relationship with area for the routing time constants is therefore conceptually acceptable.




(a)




(b)

**Figure E.7 Graphs of model fit for the slow flow routing time constant, Kb.**

### E.1.5 Quick flow routing reservoir time constant, K1

A logarithmic transformation was applied to K1 to reduce the skew of the data set. Histograms of the K1 and the logarithm of (K1+1) are presented in Figure E.8. Models constructed using the log-transformed data gave the best overall fit in terms of model fit. The best, stable model fits were obtained for a model based on HOST groupings (K1_HG). The fit obtained with the final HOST based model was similar to the fit of a model based on BFIHOST and LOG (AREA+1) but gave a better fit to low values of K1. The model is summarised in Table E.5 and plots of residuals and predicted versus observed are presented in Figure E.9a and b.

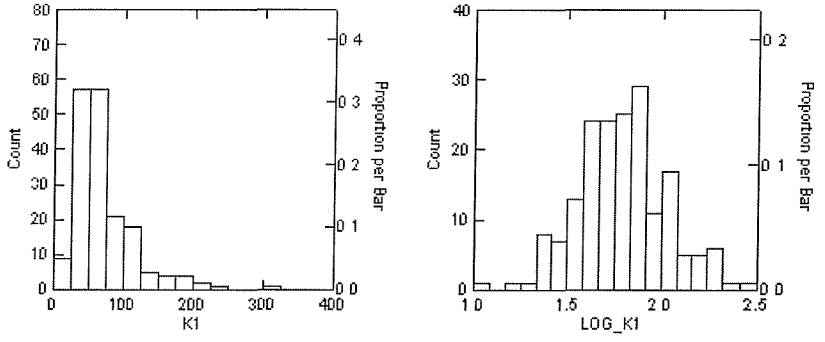


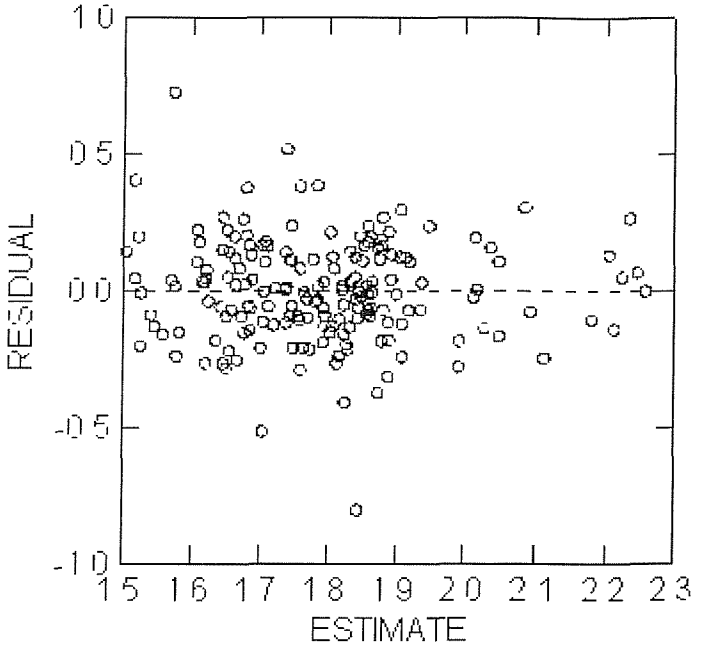

**Figure E.8 Histograms of K1 and log (K1+1).**

Two outliers were identified. The first of these was the Frome at Bishops Frome, 55028, (catchment area 77km$^2$, BFI 0.5) the observed value of 9.96 hr for K1 was much lower than the predicted value. Evidence from hydrograph showed that although the catchment has a relatively high base flows there is evidence of a quick flow response to rainfall events. The catchment is sandstone in the headwaters, which may provide a large component of the base flow whilst the lower part of the catchment is impermeable drift giving rise to a quick response component. In contrast, the second outlier, 15021 the Almond at Newtown Bridge (94km$^2$) has a large value of K1 198 hr, this is equivalent to two thirds of the calibrated value of Kb (the slow flow routing time constant) for the catchment. Visual inspection of the hydrograph showed that the catchment does not respond quickly to rainfall events justifying the large value of K1 identified through calibration. As there was no physical evidence to support the removal of these catchments they were left in the analysis.

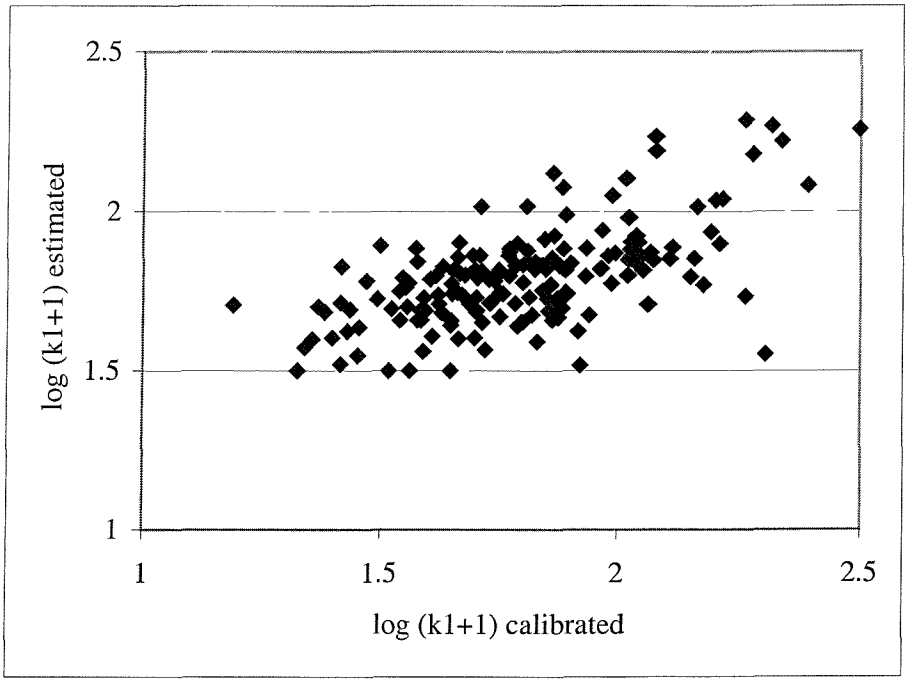
**Table E.5 Regression model for log (K1+1)**

| K1_HG | HOST | Hydrogeological units | Coefficient | Tolerance | P |
|---|---|---|---|---|---|
| 1 | 1 | Chalk, chalk drift | 0.023 | 0.949 | 0 |
| 2 | 2 | Oolitic limestone, soft magnesian | 0.022 | 0.918 | 0 |
| 3 | 5, 7 | Blown sand, gravels | 0.021 | 0.783 | 0 |
| 4 | 20, 23, 25 | Very soft massive clays | 0.016 | 0.821 | 0 |
| 5 | 17, 19, 22, 27 | Hard rock | 0.019 | 0.816 | 0 |
| 6 | 16, 18, 21 | Soft bedded clays, loams, till, shales | 0.018 | 0.882 | 0 |
| 7 | 24, 26 | Soft shales, siltstones, bedded clays/loams, clay and flints | 0.017 | 0.751 | 0 |
| 8 | 12, 28, 29 | Weathered intr/meta. Rock, raw peats, eroded peats | 0.015 | 0.716 | 0 |
| 9 | 3 | Soft sandstone | 0.018 | 0.892 | 0 |
| 10 | 4 | Weathered intr/meta. Rock, hard limestone and sandstone | 0.019 | 0.925 | 0 |
| 11 | 15,14 | Weathered intr/meta. Rock, sandstone, colluvium, loamy drift, gravels/loams | 0.015 | 0.761 | 0 |
| 12 | 11,6,13 | Earthy peats Colluvium, coverloam, gravel, loamy drift Weathered intr/meta. Rock, chalky drift, loams | 0.021 | 0.751 | 0 |
| | | adjusted $R^2$=0.99 | SE=0.195 | N=179 | |

The grouping of HOST classes is logical, based on permeability and physical charatceristics of the substrate geology. Permeable soils underlain by permeable geologies have higher coefficients. Overland flow that reaches the stream, and or minor channels will not normally be generated within very permeable systems. The direct runoff that there is will be dominated by inter-flow giving a more damped response, and hence higher values of K1. This is conceptually correct. The FSE for the model is Standard Error of 1.57, that is (given N is reasonably large) the predicitive accuracy of the model is approximately ±57% at the 68% level.




(a)




(b)

**Figure E.9 Model fits for the logarithm of the quick flow routing time constant, log (K1+1).**

The plot of residuals, Figure E.9a, and the plot of estimated values as a function of predicted values Figure E.9b, demonstrates there is a tendency to over predict very low values of K1 (including catchment 55028).

## E.2 REGRESSION MODELS FOR PREDICTING MODB PARAMETERS

The MODB parameters to be regionalised were Cmax, K1 and Kb. The loss model outliers identified in the modelling of $\overline{C}$ and Kg for MODA did not occur within the regionalisation of MODB parameters. The catchments generating the $\overline{C}$ outliers were rejected during the calibration of MODB, as the model was unable to close a satisfactory water balance. The Kg outliers did not arise as MODB uses a fixed partition coefficient to determine the partitioning of effective rainfall between quick and slow flow components. The MODA K1 outliers also did not appear as outliers in the analysis for MODB. Station 55028 was rejected, as MODB could not close a water balance. The MODA loss module is very non-linear (b=0.043 Cmax =151.76) for this catchment suggesting a potential error in the climatic data. The modelled value of K1 for 15021 with MODB was lower at 194.847 and, in the context of the other values of K1 for MODB did not constitute an outlier. The regression models for the MODB parameters are presented in the subsequent sub-sections.

### E.2.1 The maximum storage capacity, Cmax

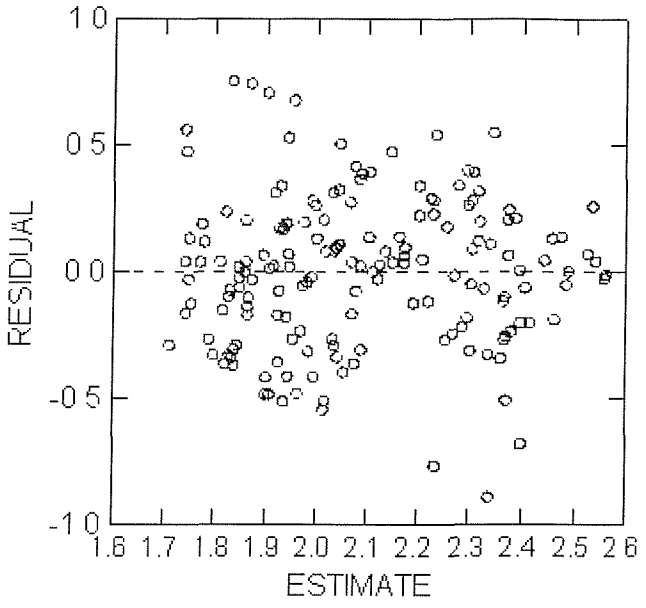
The modelling of Cmax proved to very problematical. The optimal model derived related the logarithm of (Cmax+1) to the logarithms of (PP+1) and (DPLBAR+1). The structure of the model is summarised in Table E.6 and plots of residual and estimated versus calibrated values are presented in Figure E.10a and b. The model fit is poor explaining 36% of the variance with a FSE of 1.98. The residual plot demonstrates that the model overestimates at low values of Cmax and under-estimates at high values of Cmax and therefore cannot explain the full variance within the calibrated parameter data set.

**Table E.6 Regression model for Cmax**

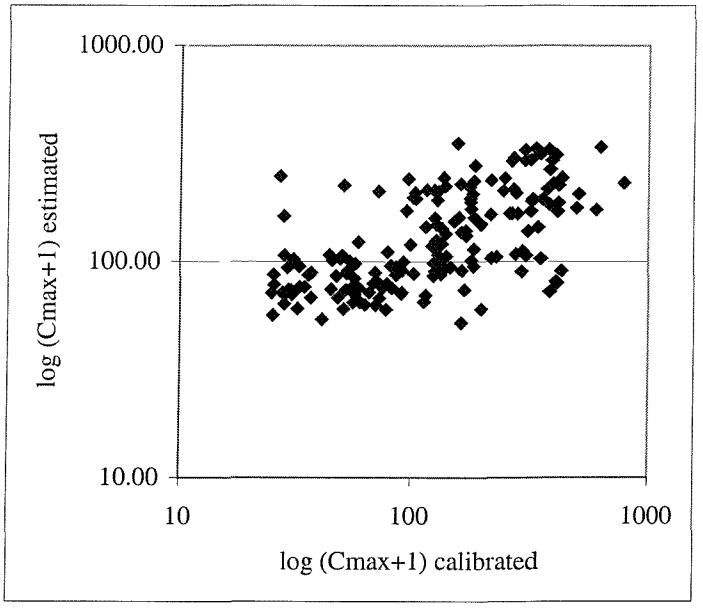
| Variable | coefficient | Std error | Tolerance | P |
|---|---|---|---|---|
| CONSTANT | 2.168 | 0.097 | | 0.000 |
| LOG_PP | 5.580 | 0.640 | 0.989 | 0.000 |
| LOG_DPLBAR | -0.261 | 0.075 | 0.989 | 0.000 |
| | Adjusted $R^2$= 0.357 | | S.E.= .296 | N=171 |

The physical interpretation of the model is not clear but may be related to the role of Cmax in controlling the evaporation rate. A uniform distribution of soil storage capacities is assumed for MODB and a fixed partition of effective rainfall is used within the model. The relationship between (precipitation-evaporative losses) and outflow from the soil store is therefore quadratic. Cmax controls the gradient of this relationship and fixes the storage capacity at which the catchment is fully saturated. Cmax also directly controls the relationship between the evaporation rate and simulated Soil Moisture Deficit (SMD). However, Cmax can only influence evaporation rates if the summer evaporation demand exceed the rainfall inputs (hence generating an SMD).

Cmax is probably only really identifiable under these circumstances. PP is a measure of the potential for significant SMDs to build up in a catchment. The regression model predicts that as PP increases so does Cmax. If PP is high the model will generate a soil moisture deficit and will potentially result in a reduced evaporation rate. Given the formulation of the evaporation function this tendency will be offset by an increased value of Cmax. The relationship between Cmax and PP may well be an artefact of the rainfall runoff model structure. The dependency on DPLBAR implies that Cmax is larger for catchments with larger mean drainage path lengths. This characteristic is high for large catchments and catchments that tend to be linear. This is difficult to explain but may be associated with the longer runoff concentration times within these catchments. The model structure is valid as it was found to be stable under the 80:20 split sample tests. The predictive capacity of the model is ±98% at a confidence level of 68%.




(a)




(b)

**Figure E.10 Graphs of model fits for log (Cmax+1).**

### E.2.2 The routing reservoir time constants, K1 and Kb

Under ideal circumstances the time constants obtained for MODA would be identical to those obtained for MODB, as they were un-correlated with the loss module parameters. However the values are influenced by the partitioning mechanism for effective rainfall. The relationships between K1 and Kb for the two model configurations are summarised in Figure E.11. This Figure demonstrates that, in general, the values obtained for the two models are very comparable. The optimal, stable regression model structures identified are

very similar to those obtained for MODA and are summarised in Table E.7 and Table E.8 respectively. The model for Kb is identical in structure to that obtained for Kb with MODA. The coefficients are different and the model is a slightly poorer fit. This may a consequence of the cruder treatment of the partitioning of effective rainfall in MODB. The model for K1 uses a log-transformed representation of K1, which is dependent on BFIHOST and Log (LDP+1). This formulation gave similar results for K1 with MODA and was only a slightly poorer fit than the HOST based model finally selected for MODA. The same formulation applied to MODB gives a predictive accuracy of ±52% which is an improvement of ± 5% over the predictive capacity of the HOST based K1 model for MODA The use of LDP gave a significant advantage over the use of AREA. Whilst these are strongly correlated they are different characteristics. The relationship indicates a damping of direct runoff response to precipitation events in long drainage path catchments. This is conceptually attractive as it represents catchment and channel routing considerations that are not explicit within the model structure.

**Table E.7 Regression model structure for Kb**

| Variable | Coefficient | Std error | Tolerance | P |
|---|---|---|---|---|
| CONSTANT | -490.9 | 123.6 | | 0.00 |
| BFIHOST | 1871.8 | 198.0 | 0.94 | 0.00 |
| LOG_AREA | 206.2 | 46.9 | 0.94 | 0.00 |
| Adjusted $R^2$ | 0.50 | | SE= 305.5 | N=171 |

**Table E.8 Regression model structure for log (K1+1)**

| Variable | Coefficient | Std error | Tolerance | P |
|---|---|---|---|---|
| CONSTANT | 0.812 | 0.09 | . | 0 |
| BFIHOST | 1.279 | 0.12 | 0.95 | 0 |
| LOG_LDP | 0.224 | 0.05 | 0.95 | 0 |
| | Adjusted $R^2$= 0.522 | | SE=0.182 | N=171 |

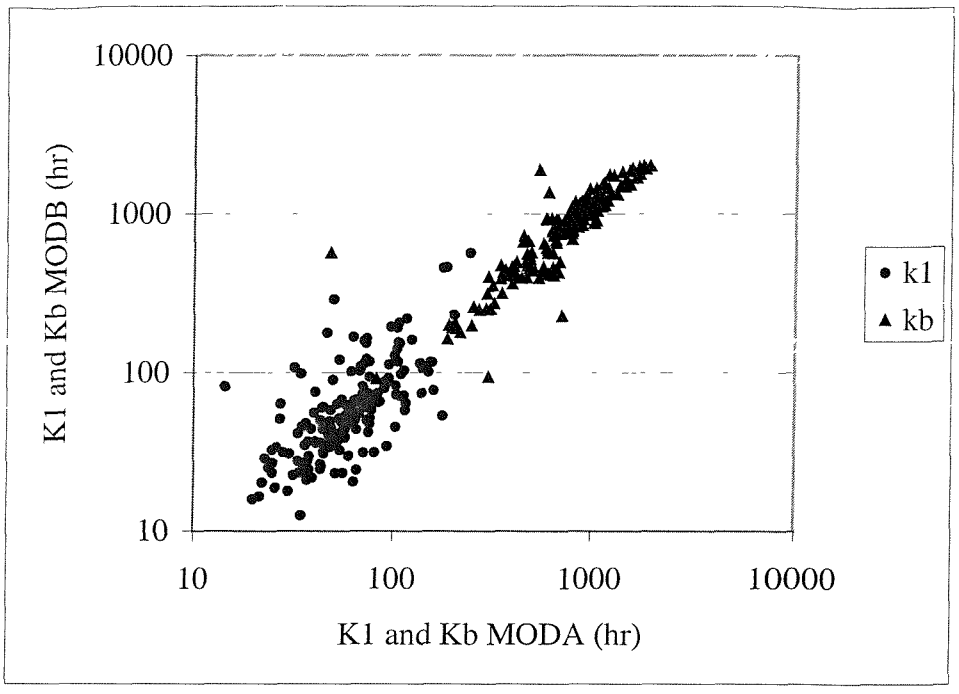



**Figure E.11 Comparison of the calibrated values of K1 and Kb parameters for MODA and MODB.**