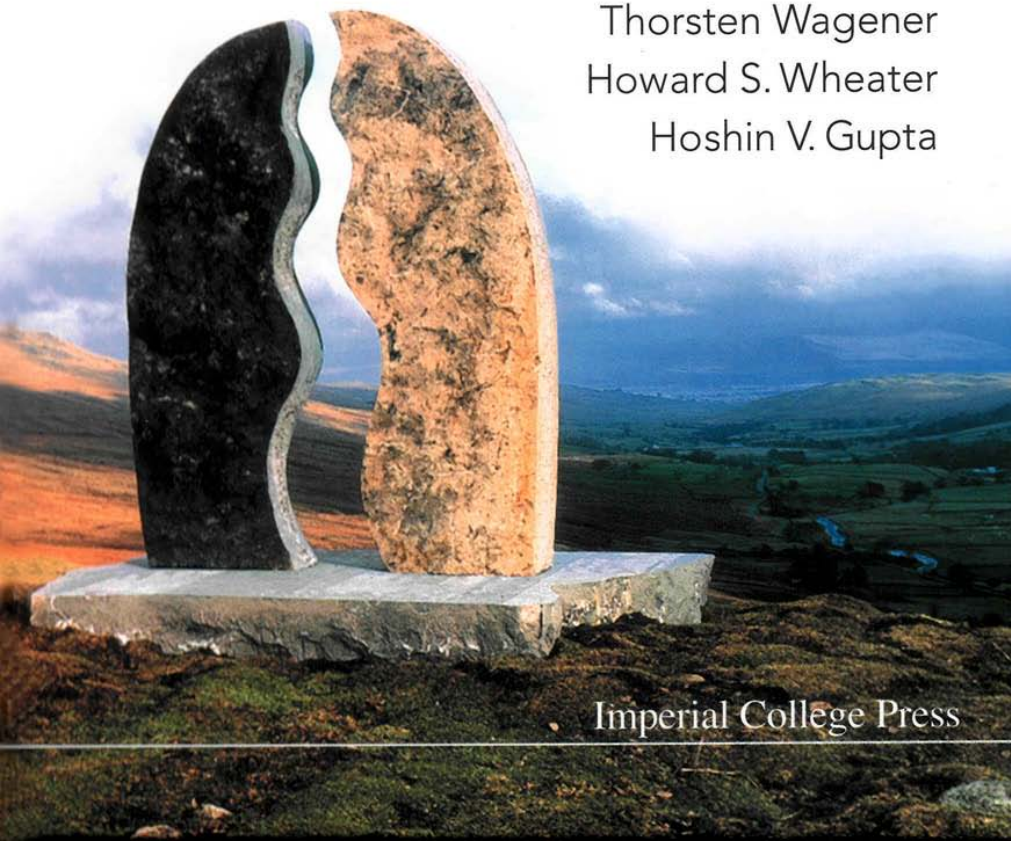
Rainfall-Runoff Modelling in Gauged and Ungauged Catchments

Thorsten Wagener
Howard S. Wheater
Hoshin V. Gupta



Imperial College Press

Rainfall-Runoff Modelling in Gauged and Ungauged Catchments

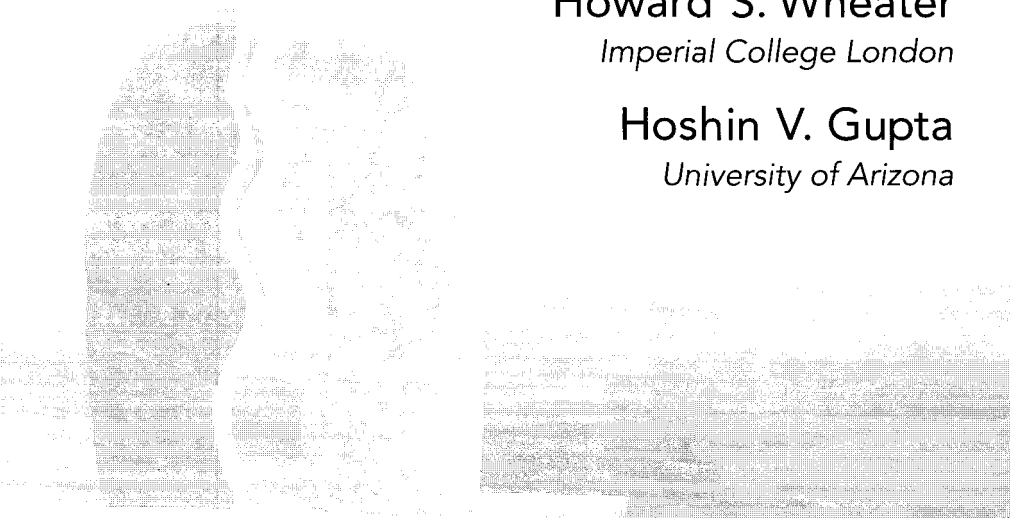
This page intentionally left blank

Rainfall-Runoff Modelling in Gauged and Ungauged Catchments

Thorsten Wagener
Pennsylvania State University

Howard S. Wheater
Imperial College London

Hoshin V. Gupta
University of Arizona



ICP
Imperial College Press

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401–402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

RAINFALL-RUNOFF MODELLING IN GAUGED AND UNGAUGED CATCHMENTS

ISBN 1-86094-466-3

Editor: Tjan Kwang Wei

Printed in Singapore by World Scientific Printers (S) Pte Ltd

für meine Eltern

– Thorsten Wagener

This page intentionally left blank

Acknowledgements

By wisdom a house is built, and through understanding it is established; through knowledge its rooms are filled with rare and beautiful treasures.

Proverbs 24, 3-4
Bible, New International Version

The work presented in this manuscript is based on the first author's Ph.D. – *Identification of parsimonious rainfall-runoff models for gauged and ungauged catchments*. A Ph.D. is very much like building a house; years of hard work, and not possible without the help of a great number of people. Thorsten Wagener would therefore like to give thanks to some important people during this project.

Such a project is usually initiated by someone who is willing to fund the house and to provide some initial ideas; thanks to Howard Wheater and Matthew Lees, especially for their continued support and constructive criticism on my changing blueprints. In addition thanks to Matthew for his contributions to the tools developed. Also thanks to the Natural Environment Research Council for grant GR3/11653, which was used to finance building the house. Thanks to the German Academic Exchange Service (DAAD) and the National Science Foundation (NSF) Science and Technology Center (STC) for Sustainability of semi-Arid Hydrology and Riparian Areas (SAHRA; EAR-9876800) for funding to paint and decorate the house.

Then there is the architect, who advises you that saddle roofs are totally out of fashion and that one should rethink what we mean by *a house* anyway. Thanks to Hoshin Gupta, for switching on some lights.

To my parents who laid the foundation, which is the hardest work of all. Your unconditional support brought me here.

To all those fellow builders (more or less experienced) who shared their thoughts with me, Neil McIntyre, Luis Camacho, Doug Boyle, Adrian Butler, Manfred Schütze, Tom Rientjes, Rob Lamb, Keith Beven, and Christian Onof. Also thanks to the contributions of some M.Sc. students, Iannis Niadas, Steven Boxall and Hyosang Lee. Thanks to Brian Greenfield from the Environment Agency, and Southern Water for providing bricks and mortar to work with. Thanks to Sue White and Christian Onof for surveying the house and judging it to be in good state and ready to be used.

To Britta and Sven, and to my friends and mates, Sven, Susi, Jo (Tigger) Geary, Luis, Helen, Neil, Jo, Nanni, Gina, Birgit, Gerrie, Luiza, Dirk, Manfred, Niels, Demetris, Modu, Alex, who gave me plants and pictures, helped me paint the rooms and made even the unfinished house a nice place to live in. Not to forget Angela and Judith.

Thanks to Mary Bourne and Keith Beven for allowing me to use their photographs of Mary Bourne's sculpture *Watercut* to decorate the walls. Thanks to Gabriella Frescura and Laurent Chaminade of Imperial College Press, London, for the excellent cooperation in getting the house on the market. Thanks to Keith Beven for writing the preface to this monograph. And, particular thanks to Mary Black for sacrificing weekends and evenings to get this house into shape.

And finally to all those that I have forgotten to mention here, you guys know me, radioactive memory ... what can you expect.

However, all labour is in vain without God's blessing and the knowledge that there is another, much more important house, already prepared for us by our Lord Jesus Christ.

In my Father's house are many rooms; if it were not so, I would have told you. I am going there to prepare a place for you. And if I go and prepare a place for you, I will come back and take you to be with me that you also may be where I am. You know the way to the place where I am going.

John 14, 2-4
Bible, New International Version

Thorsten Wagener, Tucson, Spring 2003

Preface

I have spent by far the greater part of my research career trying to do hydrological modelling, in some sense, "properly". It has been a struggle, even when trying to model experimental catchments with lots of available observational data. These academic struggles have been recorded in a plethora of papers over the last 25 years. I have not, however, been alone in finding the modelling of the hydrological response of catchments a struggle. Keith Loague, for example, in a well-known series of papers, has fought gallantly to model the R5 experimental catchment at Chickasha properly using models of ever increasing complexity and data sets of ever increasing detail. Any hydrologist who has ever tried to apply a model in earnest is only too aware of some of the problems involved.

Why is the hydrological modelling process so difficult? In part it is a result of the lack of suitable measurement techniques in hydrology. We can measure river discharges reasonably well, but the forcing function that causes those discharges is not as easy to characterize, because of the spatial and temporal variability of rainfall. Even modern radar rainfall estimation techniques do not, in most circumstances, give us sufficient accuracy, while rain gauges only provide a very small sample of the variability (even in research catchments). To model "properly", however, means more than just relating rainfall inputs to discharge outputs, it means trying to understand the flow pathways of the water within a catchment. This has long been the aim of both experimental and theoretical hydrological science, even before the days of digital computers that now allow the representation of theory as quantitative predictive models.

The development of predictive models requires understanding and representing mathematically those flow pathways under the ground surface, where the infiltrating rainfall goes. Here, our measurement techniques are even more inadequate in that they tend to give only very local information about water storage and potentials. Such measurements have allowed us to gain valuable qualitative understanding of what happens to the water as it flows towards the stream channel, or is evaporated or transpired back to the atmosphere, but the heterogeneity of soils and rocks is such that quantitative prediction in any real catchment remains very difficult.

Hydrology is also concerned with practical water management. Management requires quantitative estimates of water yields and river flows under both high and low flow conditions, not only for sites where discharge observations are available, but also for the large number of ungauged sites where there are no extant observations. In both cases, management and decision-making will require model predictions. Despite the worries of academics about the difficulties of doing modelling properly, the use of models for practical applications has become a growth industry. Indeed, the pressures for hydrological predictions are increasing as practical and sustainable water management is required to consider a wider range of variables than just water. The European Water Framework Directive, for example, will require water management to be carried out to support sustainable aquatic habitats. The resulting hydro-ecological considerations will require predictions of sediment mobilization, water quality and their impacts on important ecological criteria. It is certain, therefore, that models will be developed to meet these needs, but what should they look like?

This book presents a variety of new techniques to meet some requirements for future models of both gauged and ungauged catchments, at least for the prediction of river discharges. The techniques build on many years of experience in model development, model calibration and model application of the authors and others. They take some of the techniques, including those developed by Peter Young and myself at Lacaster which can be found in my textbook on rainfall-runoff modelling, in some novel and interesting directions.

The monograph presents an extensive study into the nature and use of conceptual rainfall-runoff models at the catchment scale. It consists of four main chapters, apart from introduction and conclusions. Chapter 2 is an extensive review of the state of the art of conceptual rainfall-runoff modelling with its particular problems regarding model identifiability. This review is followed by a chapter which details typical conceptual components found in many models and the techniques used to analyse them. Chapters 4 and 5 are studies into modelling approaches to gauged and ungauged catchments. Here underlying assumptions are tested and new approaches are introduced. Examples are the application of multi-objective techniques and a novel dynamic identifiability analysis approach.

The methods developed by the authors are presented as a set of toolboxes designed to be used by practitioners and students, or in a research context. In the toolboxes, the authors have also provided a range of uncertainty estimation techniques. This is something I have long recommended. Indeed a chapter on uncertainty estimation is something that is not commonly found. Because of my struggles in understanding the modelling process and my recognition of the difficulties of applying models to catchments with their own unique and difficult to measure characteristics, I believe that uncertainty estimation should be an intrinsic part of any hydrological modelling study. The advantage to the modeller is that giving a best estimate of the range of possible predictions is good protection against being wrong!

Of course, it is still possible to be wrong, particularly in estimating the response of ungauged catchments. This remains the ultimate problem in scientific hydrology, one that depends on trying to capture the dominant modes of response of a catchment in a parsimonious way and making the very best use of the available data to help constrain the uncertainties in such estimation. The techniques for predicting the response of ungauged catchments presented in this book for the first time are a step in the right direction.

Keith Beven, Lancaster, 2003

This page intentionally left blank

Contents

Figures

This page intentionally left blank

Tables

This page intentionally left blank

Chapter 1

Introduction

I had less difficulty in the discovery of motion of heavenly bodies in spite of their astonishing distances than in the investigation of the movement of water before our very eyes.

Galileo Galilei (1564–1642)

1.1 Rainfall-Runoff Modelling in Context

Rainfall-runoff models (RRMs) are standard tools routinely used today for hydrological investigations in engineering and environmental science. They are applied to extend streamflow time-series in space (e.g., Manley, 1978; Sefton and Howarth, 1998; Seibert, 1999b) and time (e.g., Lees and Wagener, 2000a, 2000b), to evaluate management strategies and/or catchment response to climate (e.g., Jakeman *et al.*, 1993; Cameron, 2000) and/or land use variability (e.g., Binley *et al.*, 1991; Parkin *et al.*, 1996), for the calculation of design floods (e.g., Lamb, 2000; Calver *et al.*, 1999; Cameron *et al.*, 2001), as load models linked to water quality investigations (e.g., Mroczkowski *et al.*, 1997), for real-time flood forecasting (Kitanidis and Bras, 1980a, 1980b; Moore, 1999) or to provide boundary conditions for atmospheric circulation models (e.g., Dümenil and Todini, 1992; Wood *et al.*, 1992).

A vast number of model structures, usually a combination of linear and non-linear functions, has been developed and implemented into software since the early 1960s (see Todini, 1988, for a historical review of rainfall-runoff modelling). It is therefore necessary to classify these structures and narrow down the range considered in this monograph. Probably the most commonly applied classification is one that uses three distinct classes (e.g., Wheater *et al.*, 1993). These are *metric* (also called

data-based, empirical or black-box), *parametric* (also called conceptual, explicit soil moisture accounting or grey box) and *mechanistic* (also called *physically based* or *white box*) model structures, though there might be cases where the classification of a particular model structure is ambiguous. Any parameter set within a model structure is referred to as a model in the subsequent text.

Metric models commonly use the available time-series to derive both the model structure and the corresponding parameter values. They are therefore purely based on the information retrieved from the data and do not include any prior knowledge about catchment behaviour and flow processes, hence the name black box. Among the most currently popular examples of this type of model are Artificial Neural Networks (ANN, e.g., Hsu *et al.*, 1995, 2002; Dawson and Wilby, 1998; Drecourt, 1999; Babovic and Bojkov, 2001) and Transfer Functions (TF, e.g., Young, 1984; Young, 1992). Metric models are usually spatially lumped, *i.e.,* they treat the catchment as a single unit. These models rely solely on data and are therefore, at least in theory, not suitable for the spatial extension of streamflow records into ungauged catchments. However, *data-based mechanistic* models partially overcome this problem by constraining the degrees of freedom of metric models to those structures that are physically interpretable, thereby using the hydrologist's understanding of the natural system (e.g., Young and Beven, 1994; Lees, 2000; Young, 2001).

Parametric models (O'Connell, 1991) use a storage element as the main building component. These storages (also called buckets or reservoirs) are filled through fluxes such as rainfall, infiltration or percolation, and emptied through evapotranspiration, runoff, drainage, etc. Parametric models have a structure that is specified prior to their use. In contrast to metric models, the structure is defined by the modeller's understanding of the hydrological system, and hence such models are also commonly termed *conceptual.* However, these models still rely on time-series of system output, mainly streamflow, to derive the values of their parameters in a calibration procedure. The parameters describe aspects such as the size of the storage elements or the distribution of flow between them. A number of processes are usually aggregated (in space and time) into a single parameter, which can therefore often not be derived directly from field measurements. Most of these models consider the catchment as a

single homogeneous unit. However, one not uncommon approach is the segmentation of a catchment into smaller sub-catchments, the so-called *semi-distributed* approach. Although this approach is often used, research issues related to semi-distributed approaches have only recently been defined and addressed (e.g., NWS, 2000; Boyle *et al.*, 2001; Khodatalab *et al.,* 2003). Parametric models make up the vast majority of models used in practical applications. Their dependence on flow measurements makes it difficult to apply them to ungauged catchments; however, regionalization approaches, *i.e.,* the attempt to derive statistical relationships between model parameters and catchment characteristics, have been developed which try to overcome this problem (e.g., Burn and Boorman, 1992; Abdullah and Lettenmaier, 1997; Sefton and Howarth, 1998). This has been achieved with only limited success so far. Most models used in this monograph belong to this parametric group.

Combinations of metric and parametric approaches, *i.e.,* so-called *hybrid metric-conceptual* (metric-parametric in the context of this monograph) models by Wheater *et al.* (1993), are sometimes applied (Jakeman and Hornberger, 1993). The idea is to combine the use of observations (the metric paradigm) to corroborate or reject hypothesized model structures consisting of hydrological stores (the parametric paradigm). This approach is related to the data-based mechanistic approach described earlier.

Mechanistic models (e.g., Freeze and Harlan, 1969; Beven, 1985; Abbott *et al.*, 1986; Calver, 1988; Beven, 1989; Beven, 2002) are based on the conservation of mass, momentum and energy. They became practically applicable in the 1980s, as a result of improvements in computer power. The hope was that the degree of physical realism on which these models are based would be sufficient to relate their parameters, such as soil moisture characteristics and unsaturated zone hydraulic conductivity functions for subsurface flow or friction coefficients for surface flow, to physical characteristics of the catchment (Todini, 1988), thus eliminating the need for model calibration. However, mechanistic models suffer from extreme data demand, scale-related problems (e.g., the measurement scales differ from the process and model [parameter] scales), and over-parameterization (Beven, 1989; Rientjes and Hassanizadeh, 2000). One consequence is that the model parameters cannot be derived through

measurements; mechanistic model structures therefore still require calibration, usually of a few key parameters, though applied to a large number of elements (Calver, 1988; Refsgaard, 1997; Madsen and Jacobsen, 2001). The expectation that these models could be applied to ungauged catchments has therefore not been fulfilled (Parkin *et al.*, 1996; Refsgaard and Knutsen, 1996). They are typically rather applied in a way that is similar to lumped conceptual models (Beven, 1989). Mechanistic models use a spatial discretization based on grids, hillslopes or some type of hydrologic response unit. They are therefore particularly appropriate when a high level of spatial discretization is important, e.g., to estimate local effects of soil erosion, or surface and groundwater pollution (Refsgaard and Abbott, 1996). However, if the main interest simply lies in the estimation of streamflow response at the catchment scale and if calibration data are available, then parametric models normally perform usually at least as well and the huge complexity of mechanistic models is not required (e.g., Loague and Freeze, 1985; Refsgaard and Knutsen, 1996). All models in this monograph are purely applied for streamflow prediction at the catchment scale. Mechanistic models are therefore not considered further.

1.2 Problem Analysis

The type of models considered in this monograph can be defined as *conceptual*. However, the definition applied here, after Wheater *et al.* (1993), is wider than the (probably more traditional) one described above. There is therefore a difference between those models referred to as conceptual in this monograph, and parametric models as described earlier. Conceptual models are not limited to parametric (bucket-type) models here. The conceptual classification rather includes all those models that conform to two criteria: (1) the structure of these models is specified prior to any modelling being undertaken, and (2) at least some of the model parameters do not have a direct physical interpretation, in the sense of being independently measurable, and have to be estimated through calibration against observed data.

The task of the modeller is to identify a model, *i.e.*, a parameter set

within a specific model structure, that is suitable for the given catchment characteristics, data and modelling purpose (Wagener, 1998). The identifiability process therefore consists of two steps (Sorooshian and Gupta, 1985): the identification (selection) of an appropriate model structure, and the identification (estimation) of an appropriate parameter set within this structure. This task is difficult due to a range of uncertainties involved in the modelling process that are also unavoidably propagated into the model output. These uncertainties stem from different sources (e.g., Kitanisis and Bras, 1980a; Melching, 1995; Høybye, 1996; Bastidas *et al.*, 1999), mainly:

- *Data uncertainty*, *i.e.,* uncertainty introduced by errors in the measurement itself or by data pre-processing.
- *Model specification uncertainty*, *i.e.,* the inability to converge to a single 'best' parameter set (model) using the information provided by the available data. This is often referred to as the identifiability problem.
- *Model structural uncertainty* introduced through simplifications and/or inadequacies in the description of real world processes.

Additionally, one has to keep in mind that even if those uncertainties could be removed, there would still be some (unmeasurable) randomness in the natural processes themselves (Melching *et al.*, 1990). This randomness introduces uncertainty that cannot be reduced.

Many researchers have found that they were unable to locate a meaningful global optimum in the feasible parameter space for different model structures (e.g., Johnston and Pilgrim, 1976; Gan and Burges, 1990; Sorooshian *et al.*, 1993; Kuczera, 1997), where the feasible space is defined by the ranges in which possible values of individual parameters can lie. It was initially thought that this was simply a matter of developing more powerful automatic search algorithms (e.g., Duan *et al.*, 1992). However, these are available now (see the comparison studies by Gan and Biftu, 1996; Kuczera, 1997; or Thyer *et al.*, 1999), but the problem still exists.

Consequently, some researchers abandoned the idea of uniqueness or point identifiability of a parameter set. They replaced it by an approach based on interval or regional identifiability of possible parameter sets (Spear and Hornberger, 1980; Beven and Binley, 1992). Different parameter sets within such a region yield similar results in terms of a par-

ticular objective function, often called *behavioural* parameter sets (Spear and Hornberger, 1980), thus reflecting the fact that their performance is consistent with observed behaviour.

The problem of identifiability is further increased when model structural uncertainties are considered (e.g., Gupta *et al.*, 1998). Due to the inadequacy of many model structures, one parameter set may be insufficient to represent the behaviour of the catchment during different response modes such as low and high flow periods. A consequence of these uncertainties in the case where sufficient streamflow data are available for model calibration (*i.e.*, the gauged catchment case) is that predictions have to be given for every time step as a range instead of single values (e.g., Beven and Binley, 1992; Uhlenbrock *et al.*, 1999).

The result of this lack of identifiability, with respect to both model structure and parameters, is especially severe when runoff estimates in ungauged catchments are required. *Regionalization* of model parameters has long been attempted as a methodology to circumvent the problem of the absence of calibration data in ungauged catchments (Mainley, 1978; Sefton *et al.*, 1995; Abdulla and Lettenmaier, 1997; Sefton and Howarth, 1998; Post *et al.*, 1998; Xu and Singh, 1998; Vogel *et al.*, 1999). The basic idea is to calibrate a specific model structure to as many catchments as possible and to derive statistical relationships between model parameters and catchment characteristics. These relationships can then be used in *hydrologically similar* ungauged catchments to derive parameter estimates. This has been done with some success for event-based model structures that usually only contain a couple of parameters (e.g., NERC, 1975; Dyer *et al.*, 1994). The success for more complex continuous model structures however has been very limited. Continuous modelling of ungauged catchments is therefore an unsolved problem.

In summary: The lack of identifiability of model structure and parameters is a major restriction of current conceptual RRM applications. It introduces large uncertainties into model predictions and seriously limits the possibility of sensible parameter regionalization for the modelling of ungauged catchments.

1.3 Conclusions and Resulting Research Objectives

Conclusions drawn so far can be summarized as follows:

- Lumped conceptual models perform as well as distributed complex models in cases where calibration data are available and where the modelling purpose is purely to forecast flow at the catchment outlet.
- Even lumped conceptual models with relatively simple structures are often hampered by a lack of structural and parameter identifiability.
- No completely successful application of these simple continuous models to ungauged catchments has been reported yet. It is suggested that this is mainly a consequence of lack of identifiability.

The following research objectives result from these initial conclusions:

- To build a toolkit for development, application and evaluation of parsimonious lumped continuous RRMs in line with the state-of-the-art.
- To apply this toolkit to the local modelling of gauged catchments and improve on existing procedures to evaluate the suitability of specific and competing model structures and models respectively in order to reduce the problem of lack of identifiability.
- To review the regional application of current conceptual models to ungauged catchments by means of parameter regionalization and devise an improved approach considering, amongst other things, the results of the investigation of local modelling.

1.4 Organization of Monograph

The monograph is structured into seven chapters. Each chapter builds on the results and conclusions of the previous chapters. However, the basic assumptions underlying the work presented here are re-iterated at the beginning of each chapter. This allows for chapters to be considered to be 'stand-alone'.

Chapter 1 is an introductory chapter, giving a short overview of rainfall-runoff modelling in the context of this monograph. It also analyses

the problems to be tackled and defines the objectives to be reached.

Chapter 2 is a review of the state-of-the-art of parsimonious rainfall-runoff modelling in gauged and ungauged catchments. It contains a definition of the principle of parsimony and how it can be applied to hydrological modelling, a description of typical structural components of lumped parsimonious models, and a review of modelling procedures for gauged and ungauged catchments.

Chapter 3 is a description of a toolkit for development, application and evaluation of lumped and parsimonious rainfall-runoff models (Wagener *et al.*, 2002a), which was designed and implemented in the course of this research.

Chapter 4 analyses conceptual modelling in gauged catchments, here referred to as local modelling. The current paradigm of RRM calibration is analysed followed by the introduction of a multi-objective framework for the evaluation of competing local model structures and models (Wagener *et al.*, 2001). A new approach to the evaluation of individual conceptual model structures and models, *i.e.*, parameter and model structure combinations, called DYNamic Identifiability Analysis (DYNIA) is presented (Wagener *et al.*, 2003a). These two approaches are combined in a procedure for rainfall-runoff model development and application (Wagener *et al.*, 2003b; Wagener and Wheater, 2002).

Chapter 5 is an investigation into the modelling of ungauged catchments, called regional modelling in the context of this monograph. The chapter analyses the theory behind this approach, various regional model structures and regional modelling procedures. The results are combined in a framework for regional modelling.

Chapter 6 concludes the monograph with discussion and conclusions derived from the research presented here. Recommendations for future research are made.

Appendix A shows some early investigations into regionalization and the effects of hydrograph segmentation on parameter sensitivity (Wagener *et al.*, 1999). *Appendix B* and *Appendix C* describe briefly how the two Matlab toolboxes – the Rainfall-Runoff Modelling Toolbox (RRMT) and the Monte Carlo Analysis Toolbox (MCAT) – can be obtained and installed.

Chapter 2

Rainfall-Runoff Modelling – A Review

All our knowledge has its origins in our perception.

Leonardo da Vinci (1452–1519)

2.1 Introduction

This chapter presents a review of the nature and use of lumped parsimonious models for the continuous simulation of the rainfall-runoff (RR) relationship. The description of the hydrological processes occurring in the catchment is restricted to circumstances similar to those found in the test catchments used in Chapters 4 and 5, *i.e.,* humid areas. Detailed algorithms of approaches used are given in Chapter 3.

2.2 The Principle of Parsimony in Hydrological Modelling

Different researchers have observed that the number of parameters required to describe the key behaviour of environmental systems is often quite low (e.g., Jakeman and Hornberger, 1993; Young *et al.*, 1996). Increasing the degree of model complexity, *i.e.,* the number of free parameters, above a certain level does not result in a significantly improved model performance in terms of a better reproduction of the catchment behaviour (e.g., Naef, 1981; Hornberger *et al.*, 1985; Refsgaard and Knutsen, 1996). Instead, the problem arises that many parameter combinations, often widely distributed over their individual feasible ranges, lead to acceptable model performances (e.g., Spear, 1995; Kuczera and

Mroczkowski, 1998), an effect sometimes called *equifinality* (Beven, 2000a, p.244).

It is often possible to analyse the necessity of the model parameters to reproduce the catchment behaviour and reduce the model complexity to a (minimally) required level using sensitivity analysis. Wagener *et al.* (1999; see Appendix A) reduced the number of free parameters of an Antecedent Precipitation Index-based model structure from seven to three to reproduce the (monthly) streamflow behaviour of 23 catchments in the USA. This resulted in much increased identifiability of the remaining parameters, while the model performance decreased only slightly.

Results from other research efforts in the field of rainfall-runoff modelling suggest that usually up to five or six parameters can be identified from daily streamflow and rainfall data using traditional single-objective calibration schemes (e.g., Kirkby, 1975; Hornberger *et al.*, 1985; Wheater *et al.*, 1986; Beven, 1989; Jakeman and Hornberger, 1993; Ye *et al.*, 1997; Gaume *et al.*, 1998; Perrin *et al.*, 2000). Some researchers have therefore concluded that only models with no more than half a dozen parameters are required to describe the behaviour of a catchment with respect to the production of daily streamflow (e.g., Beven, 1989; Jakeman and Hornberger, 1993; Beck, 1987). These findings have led to the investigation of less complex (often termed parsimonious) model structures that only capture the key response modes of the hydrological system (e.g., Hornberger *et al.*, 1985; Jakeman and Hornberger, 1993; Young *et al.*, 1996).

The *principle of parsimony* requires models to have the simplest parameterization that can be used to represent the data (Box and Jenkins, 1976). Parsimonious model structures have reduced problems of identifiability since only model parameters justified by the data are kept. The principle of parsimony, also known as Ockham's Razor, was advocated by the Franciscan monk William of Ockham in the early 14th century in statements such as "plurality should not be posited without necessity," or "what can happen through fewer [principles] happens in vain through more" (Spade, 2000, p.101). The principle can be described as statement of a *cautious theoretical method* (Spade, 2000, p.102). The approach of retaining only necessary components ensures that the model components used are positively affirmed. However, using this approach in the context

of RR modelling does not guarantee that all necessary model components are identified. Careful consideration is therefore required to ensure that the model does not omit one or more hydrologic processes important for a particular problem. A model structure that is too simplistic in terms of the number of processes represented can be unreliable outside the range of catchment conditions (*i.e.,* climate and land use) on which it was calibrated (Kuczera and Mroczkowski, 1998). It is therefore vital to use data with high information content in order to ensure that the main response modes are excited during calibration (Gupta and Sorooshian, 1985a; Yapo *et al.*, 1996). Young *et al.* (1996) describe how the principle of Ockham's razor can be applied to the modelling of environmental systems using a combination of Monte Carlo techniques, dominant mode analysis, and data-based mechanistic modelling.

Popper (2000, p.142) advocated the use of simple theories based on their degree of testability. The justification is that simpler theories apply more restrictions on how a system is allowed to behave and therefore are easier to test than complex theories. System behaviour not permitted by the theory will lead to rejection or modification of the theory. The degree of testability of a theory is therefore proportional to the amount of behaviour prohibited by it (Popper, 2000, p.119). In the context of hydrological modelling, it is important to consider that the testability of a model structure will be increased in cases where additional output variables exist that can be compared to observations, e.g., predictions of groundwater levels or saturated areas (Nash and Sutcliffe, 1970; Seibert, 1999a). Additional information is then available to test potential models, and hence to reject those that show behaviour that does not conform with observations.

2.3 The Rainfall-Runoff Process

The hydrological processes occurring in and above a catchment, from the formation of rainfall to the streamflow that finally leaves the catchment through a river, are many and complex (Fig. 2.1). The most important of them, with respect to rainfall-runoff modelling, are briefly described here. Emphasis is put on processes that occur in humid areas, while snow

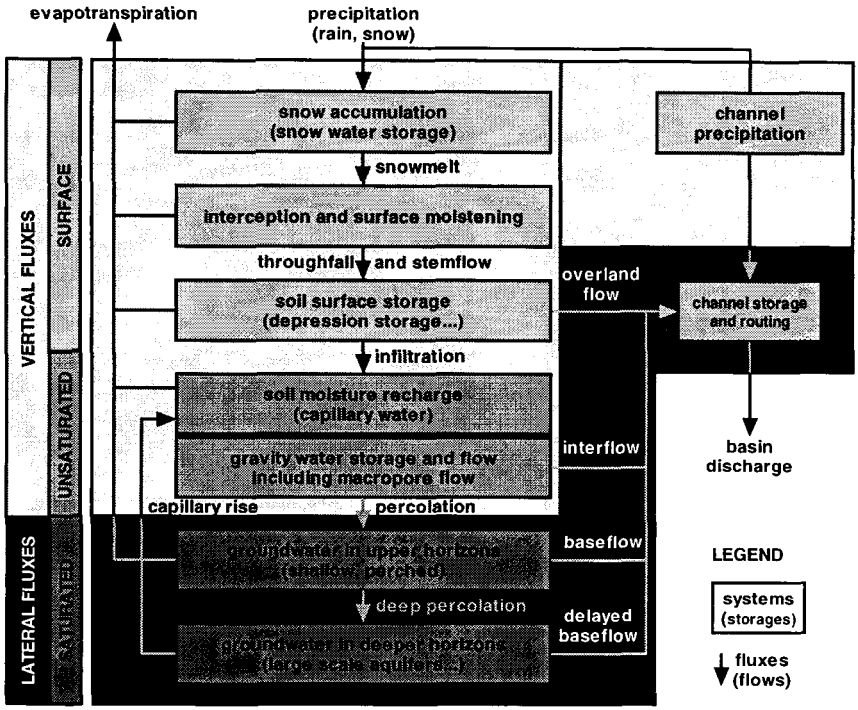


Fig. 2.1 Schematic presentation of hydrologic processes in a catchment modified from Becker (1992).

formation and melt is omitted since it is of no particular importance for the test catchments considered later in this monograph.

Precipitation occurs when water condenses around aerosol particles (such as dust or smoke), driven by the cooling of air masses through upward movement. In humid environments such as the UK, this upward movement is usually caused by warm moist air being lifted through colder denser air moving underneath it on a large scale, *i.e.*, frontal precipitation, but can also be caused by air masses being warmed by heat originating from the ground surface, *i.e.*, convectional precipitation. The total amount of precipitation is often referred to as *gross precipitation*.

A small percentage of the precipitation falls directly onto the stream network and thus contributes immediately to runoff as *channel precipitation*.

Part of the precipitation (approximately 1 to 1.5 mm) is intercepted by the vegetation canopy, from which much evaporates back into the air, the so-called *interception* loss (Ward and Robinson, 2000, p. 72). The rest forms the *net precipitation*; it moves down the vegetation as stemflow, drips off the leaves, or directly falls to the ground as throughfall.

Most of the intercepted precipitation and part of the water reaching the ground or even infiltrating into the upper soil layer(s) returns to the atmosphere through *evapotranspiration*. This is the combined effect of evaporation of water directly from soil or open water, and of transpiration by the vegetation. Evaporation involves a change of state, from liquid water to water vapour, which requires an input of energy (latent heat). This energy is mainly provided by solar radiation.

The remaining *net precipitation* fills surface depressions, infiltrates into the soil and sometimes even directly forms *surface runoff*. Surface runoff can be caused by an *infiltration excess mechanism* where the rainfall intensity is higher than the infiltration capacity of the soil, or by a *saturation mechanism* where the soil is saturated and cannot take up further water (Dunne, 1978; Ward and Robinson, 2000, p.239ff.). The former situation is unlikely to occur on soil with a dense vegetation cover, which increases the infiltration capacity of the soil considerably at least for temperate climates. This leaves the latter mechanism as the main cause for overland flow in humid environments such as the UK. These saturated (or riparian) areas occur mainly adjacent to streams and on lower valley slopes. They are dynamic, *i.e.*, they grow during a rainfall event, and are therefore often referred to as variable (dynamic) contributing or source areas (Beven, 2000a, p.11). Topography has a dominant role in controlling these areas (Hornberger *et al.*, 1998, p.209). Saturated areas form at locations where large upslope areas drain (e.g., convergence of flow at hillslopes or hollows) and where the capacity for downslope drainage is limited (where the slopes flatten towards the stream) (Hornberger *et al.*, 1998, p.211). The surface runoff itself is more likely to occur in small rivulets than in the often-assumed sheet flow (Dunne, 1978; Beven, 2000a, p.11).

The subsurface can often be divided into two overlying zones, a zone of aeration and a saturated zone. The capillary fringe forms the boundary

between those two zones, while the former can be further subdivided into a soil and an intermediate zone.

The *soil zone* is of major importance for runoff processes because water within it can be divided into storage, throughflow, percolation and water losses through evapotranspiration. Soil moisture storage is commonly simplified as follows: The available water capacity for the vegetation rooted in the soil zone is the amount between *wilting point* (WP) and *field capacity* (FC). Below WP, the plant is no longer able to overcome the absorption forces binding the water to the soil, and above FC, the water will drain from the soil zone due to gravitational forces. The water taken up will be used by the plant or given back to the atmosphere through transpiration. *Throughflow* (also called *interflow*), lateral flow through the soil zone, occurs as shallow perched saturated flow in cases where the lateral hydraulic conductivity is higher than the vertical one, and a sufficiently high slope is present. Other contributors to throughflow are the flow through macropores (explained below) or, less likely, unsaturated flow through the soil profile. Throughflow can be responsible for up to 85% of the total runoff outside arid and semi-arid areas (Ward and Robinson, 2000, p.238).

Water percolating from the upper soil enters the *intermediate zone* and moves further down to the *capillary fringe*. The latter is a zone above the water table in which water is held by capillary forces to fill most of the soil pores. This (dynamic) zone can extend into the soil zone in the floodplains of a catchment, while a distinct intermediate zone is commonly found at the valley flanks.

Below this lies the saturated zone, the *groundwater*. If the soil or rock that contains the groundwater is capable of transmission of significant quantities of water, it is called an *aquifer*, which may be *confined* or *unconfined*. The former has an overlying less permeable layer, while the latter's upper boundary is the water table. Hydrologically, aquifers commonly contain confined and unconfined areas as part of one unified system (Ward and Robinson, 2000, p. 143). The groundwater table is usually connected to the river through the river flanks. The above-mentioned topographic control on water movement can be extended to the subsurface if catchment characteristics allow the assumption *that the hydraulic gradient of the saturated zone can be approximated by the local surface*

topographic slope measured with respect to the plan distance (Beven, 2000a, p.188).

Rapid, preferential movement of water in the soil can also occur through *macropores*, large openings or voids caused for example by structural cracks, earthworms or decay of plant roots. This type of flow can result in water bypassing the upper soil, yielding quick recharge of groundwater, or adding to the throughflow. *Macropore flow tends to dominate soil water flow processes during, and shortly after, rainfall but becomes less important than matrix flow during the subsequent period of redistribution* (Ward and Robinson, 2000, p.225).

Channel precipitation, surface runoff and part of the throughflow form the quick (direct) contribution to runoff, while the remaining throughflow and the groundwater flow constitute the slow (delayed) contribution to runoff. Direct and delayed runoff together result in the streamflow measured at the catchment outlet, also referred to as discharge or catchment yield.

Detailed descriptions of hydrological processes can be found in the textbook by Ward and Robinson (2000) or the texts by Kirkby (1985), Troendle (1985), or Dunne (1978).

2.4 Structural Representation in Parsimonious Models

The first continuous and lumped (parametric) RR models, developed in the 1960s and 1970s, were often direct translations of the hydrologist's *perceptual model* (Beven, 2000a, p.6) into mathematical structures. The idea was to make these model structures, and thus their parameters, as physically realistic as possible, without consideration of aspects such as parameter identifiability. *The ideal model would specify completely the properties and processes that occur in all the relevant components of the basin. The specification would be given in terms of physical parameters and would involve all the behavioural relationships within the basin* (Dawdy and O'Donnell, 1965). These structures show a high level of complexity especially with respect to soil moisture processes, *i.e.*, the mainly vertical fluxes such as infiltration, percolation, etc. Examples of these structures are the Stanford Watershed Model (16 parameters, Craw-

ford and Linsley, 1966; or its Fortran version the Kentucky model; James, 1973), the Sacramento Soil Moisture Accounting Model (SAC-SMA, 16 parameters, Burnash, 1995), the Hydrological Simulation Model (HYSIM, 17 parameters, Manley, 1978), the Monash Model (13 parameters, Porter and McMahon, 1971), and the Nedbør Afstrømnings Model (NAM, 14 parameters, Nielsen and Hansen, 1971). The number of (free) parameters might vary between different versions of these model structures.

However, it has become apparent that these model structures can suffer from over-parameterization, *i.e.*, the structures contain more parameters than can be identified in a calibration procedure, especially after the introduction of automatic optimization procedures to parametric RR model structure parameter estimation (Dawdy and O'Donnell, 1965).

Ibbitt (1970) and Pickup (1977) report problems of estimating the *true* parameters of the Stanford watershed and the Boughton model structure respectively using automated procedures and synthetic error free data. Johnston and Pilgrim (1976) were unable to find a unique *best* parameter set (with respect to the sum of squared residuals) for the (9 parameter) Boughton model structure in two years of work utilizing real data from an Australian catchment.

These problems have led to comparison studies using model structures with different levels of complexity in order to identify the levels actually supported by the information content in hydrological time-series (e.g., Naef, 1981; Hornberger *et al.*, 1985; Jakeman and Hornberger, 1993; Michaud and Sorooshian, 1994; Gan *et al.*, 1997; Uhlenbrock *et al.*, 1999; Perrin *et al.*, 2000).

A model comparison study (10 model structures on 6 different catchments, semi-arid to humid, 1580 to 104000km^2) conducted by the World Meteorological Organization (WMO, 1975) resulted in the general conclusion that simple models perform as well as complex ones in humid catchments, while the selection of a suitable structure for semi-arid conditions is more critical. Based on the results of this and a second study (Naef, 1977), Naef (1981) concluded that *it could not be proved that complex models give better results than simple ones.* Many structures, even simple ones, will give good results during average conditions. However, the studies suggest that all models will perform poorly when it

comes to forecasting under extreme conditions or during extrapolation. The assumed reasons are structural shortcomings in all models and the lack of good quality data.

Hornberger *et al.* (1985) used a 13-parameter version of Topmodel (Beven *et al.*, 1995) to model a catchment (5.15km^2) in the USA. They found that parameter sets giving a good performance were widely distributed in the feasible space. Hornberger *et al.* (1985) subsequently developed a reduced version of the initial, complex model through elimination of those model components that seem to be less important in the specific hydrological system under investigation, based on sensitivity analysis. The lack of identifiability improved considerably when the model complexity was reduced. The new structure showed an almost equally good fit compared to the full model structure.

Franchini and Pacciani (1991) compared seven conceptual model structures (Stanford IV, Sacramento, Tank, APIC, SSARR, Xinanjiang, Arno) using hourly data from an Italian catchment. All model structures were split, as far as possible, into a water balance and a routing component. An approach was made to unify the routing component (by adopting the Arno approach) and the method to calculate the potential evapotranspiration of all structures. The comparison therefore concentrated on the water balance component. Franchini and Pachiani's (1991) summarized result was that *significantly different models produced equivalent results, with calibration times generally proportional to the complexity* (*i.e.,* number of parameters) *of their structure.*

In contrast to Franchini and Pachiani (1991), Jakeman and Hornberger (1993) studied the identifiability of (linear transfer function) routing structures (in connection with a non-linear loss function) from rainfall, temperature and streamflow time-series, using catchments ranging from 0.00049 to 89.6km^2 and from semi-arid to humid climates. They concluded that a two-component (parallel) structure with four parameters was sufficient for routing when modelling at a daily time-scale. One component could be related to a quick catchment response and one to a slow catchment response, without further specification of flow paths. Only one component (two parameters) could be identified if either baseflow was absent (see also Ye *et al.*, 1997; 1998), *i.e.,* for ephemeral streams, or the time scale was coarser than daily.

Investigations by Gan *et al.* (1997) showed little difference in performance when applying a range of model structures (6 to 15 free parameters) to three African and USA catchments (2344, 2480 and 2682 km^2). Their result was that it is not the degree of complexity (the number of parameters), but the type of process description which made a difference. It was also shown that *on the whole, dry catchments are more sensitive to the model structure and harder to model than wet catchments* (Gan *et al.*, 1997). The ability of one model structure (Xinanjiang or XNJ; Zhao and Liu, 1995) to account for the non-uniform distribution of runoff producing areas was assumed to be responsible for a better performance in dry catchments.

Uhlenbrock *et al.* (1999) found that they could achieve good predictions with different parameterizations (*i.e.*, different model structures) of the HBV model framework (Bergstrom, 1995) applied to a mountainous German catchment (39.9 km^2). All structures showed identifiability problems with respect to their parameters, and, of more concern, they found it impossible to refute (known) incorrect representations of the natural system based on performance measures.

Perrin (2000; see also Perrin *et al.*, 2000) compared 19 continuous lumped models operating on a daily time scale on 429 catchments in France (307), USA (82), Australia (26), the Ivory Coast (10) and Brazil (4). These catchments varied widely in characteristics such as size (0.1–50600 km^2), mean annual precipitation (300–2300 mm) or baseflow index (0.1–98.5%). The general conclusion of his research was that complex model structures outperform simple ones in the calibration stage of a split-sample test, but not during validation on an independent data period. He related this lack of robustness to model over-parameterization.

The overall result of these studies is that, when sufficient calibration data are available, simple structures (commonly) perform as well as complex ones (at least when combining calibration and simulation performances) under humid (wet) conditions. However, simple models reduce the problem of lack of identifiability considerably and three to five parameters seem sufficient to represent the (daily) streamflow behaviour of a catchment. A simple routing component, which allows the split into a quick and a slow contribution, appears sufficient for continuous modelling on a daily time-scale.

Structures of this level of complexity are subsequently referred to as parsimonious. A range of parsimonious structures is available, e.g., IHACRES (the Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Streamflow data; Jakeman *et al.*, 1990; Jakeman and Hornberger, 1993), the Probability Distributed Model (PDM; Moore, 1999), GR3J (the modèle du Génie Rural à 3 (4,5) paramètres Journalier; Edijatno *et al.*, 1999; Perrin, 2000), the Time Area Topographic Extension (TATE; Calver, 1993; 1996) or the ABCD (Alley, 1984) model structures, though the range of components used in these structures is limited. The following overview therefore describes common components, rather than specific model structures.

For this purpose, it is convenient to separate the hydrological processes occurring in the catchment into two domains, following a suggestion by Becker and Nemec (1987, Fig. 2.1). They segment the catchment into a domain of mainly vertical fluxes, more related to the water balance or soil moisture accounting (SMA) part of the model structure, and a domain of mainly lateral fluxes or flows, more related to the water transport or routing component(s). The SMA component uses rainfall and potential evapotranspiration (PE) or temperature as inputs and separates the rainfall into losses through actual evapotranspiration (AE), storage, and the part contributing to runoff, usually referred to as effective rainfall (ER). A routing or water transport component represents the lateral flows domain by adding translation effects to the ER to produce streamflow. Possible approaches to represent these components are described below. (A detailed mathematical description of selected components can be found in Chapter 3).

2.4.1 Soil moisture accounting

The oldest approach to estimate runoff from a catchment is probably the rational formula, dating back to the mid-19th century. It is an event-based approach relating rainfall and catchment area to runoff, using a catchment-specific coefficient. The formula takes the form

$$Q = C \cdot r \cdot A \tag{2.1}$$

where Q is the streamflow at the catchment outlet, A is the catchment area, r is the rainfall, and C is the coefficient (e.g., Chow *et al.*, 1988, p.496ff.). One assumption in this approach is that C is a time-independent coefficient. In reality it will depend, among other aspects, on the antecedent moisture condition in the catchment and is often used with other implicit assumptions, e.g., to relate rainfall of a given frequency with runoff of the same frequency.

A simple representation of antecedent conditions is the widely used Antecedent Precipitation Index (API, e.g., Shaw, 1994, p.335),

$$API_t = k \cdot API_{t-1} + r_{t-1} \tag{2.2}$$

where API_t is the index t days after an initial starting value. Parameter k accounts (mainly) for evapotranspiration and drainage losses over time (typical values lie between 0.85 and 0.98). This equation assumes an exponential decay and is calculated using daily time-steps. This empirical approach has been adopted for the calculation of effective rainfall, for example in the IHACRES model structure (Jakeman *et al.*, 1990; Jakeman *et al.*, 1994; see Chapter 3).

A more physically realistic representation of the water balance became possible after Penman (1948) and Thornthwaite (1948) developed methods to estimate actual evapotranspiration from meteorological data and catchment (vegetation and soil) characteristics (Blackie and Eeles, 1985, p.314). The water balance could now be stated explicitly as

$$Q(t) = r(t) - ae(t) - \Delta S(t) \tag{2.3}$$

where Q [L] is the streamflow, r [L] is rainfall, ae [L] is the actual evapotranspiration and ΔS [L] is the change in moisture storage in the catchment. This water balance (in explicit or implicit form) is the basis of most parametric approaches to SMA.

The simplest approach uses a simple store (bucket), representing the storage capacity of the catchment (e.g., Manabe, 1960). Water is added through precipitation and lost through evapotranspiration. Effective rainfall is produced through overflow of the store. This accounts for the production of runoff after large events, when the catchment is sufficiently wet. It does not account, however, for the fact that the catchment often drains long into dry periods (Evans and Jakeman, 1998). This can par-

tially be accounted for through the routing component, but is probably more realistically addressed through the addition of a bottom outlet to the store. Evans and Jakeman (1998) apply this approach in their catchment moisture deficit (cmd) component.

As described earlier, evapotranspiration will occur at a high rate from the soil while sufficient water is available, but at a much-reduced rate as soil moisture is depleted. One representation of this is a structure consisting of different stores in a vertical setting, each with a different AE rate (Calver *et al.*, 1993). Penman (1949) developed a model that can be represented as a two-layer structure from which evapotranspiration occurs at potential rate from the top store, while it is reduced to one-twelfth of that rate for the bottom store. The layers are connected through an overflow mechanism.

Actual evapotranspiration, based on measurements of temperature or estimates of PE, is often described using simple empirical functions (e.g., as used in IHACRES) or as a functional relationship to the actual soil moisture content (e.g., PDM, HBV, Penman). This relationship can be linear (e.g., PDM), linear with thresholds (e.g., Penman, HBV), or quadratic (e.g., PDM).

A quick contribution of rainfall to runoff or to deep percolation, e.g., through infiltration excess overland flow or the effect of macropores, is often present. This effect can be introduced in form of a constant fraction of rainfall bypassing the soil moisture stores (Penman, 1949; Thornthwaite and Mather, 1955; Alley, 1984; Jolley, 1995, p.402).

However, these quick contributions are dynamic and not constant in time, as assumed with the bypass approach. This is considered in the Hydrologiska Byråns Vattenblanasavdelining (HBV; Bergström, 1995; Harlin, 1991; Seibert, 1997) model structure, where the production of effective rainfall is exponentially related to the actual content of a soil moisture component.

A parallel distribution of stores with different capacities can be used to account for the fact that different areas within a catchment will respond differently to rainfall input. For example, areas in valley bottoms or adjacent to rivers might become saturated relatively quickly, while areas higher up the hillslope might never reach saturation and therefore do not contribute to quick runoff. The idea of a probability distribution of

storage capacities is described in detail by Moore and Clarke (1981; Moore, 1985, see Chapter 3), but can also be found in many other model structures, e.g., PDM (Moore, 1999), Xinanjiang (XNJ, Zhao and Liu, 1995), Variable Infiltration Capacity (VIC; Wood *et al.*, 1992; Abdulla and Lettenmaier, 1997), and Arno (Dümenil and Todini, 1992). This approach considers the heterogeneity of the catchment response without allowing the heterogeneity to be mapped onto specific locations in the catchment.

This mapping is possible within the Topmodel approach (Beven and Kirkby, 1979; Beven *et al.*, 1995). Topmodel is based on the earlier described topographic control on water transport. It is derived from two assumptions (Beven, 2000a, p.188ff.): (1) the dynamics of the saturated zone can be approximated by successive steady-state representations of the saturated zone on an area a draining to a point on a hillslope; and (2) the hydraulic gradient of the saturated zone can be approximated by the local surface topographic slope measured with respect to plan distance $\tan\beta$. Using these topographical characteristics in the form of an index $a/\tan\beta$, and adding a third assumption, with respect to the distribution of the lateral saturated transmissivity, leads to an index of hydrological similarity, *i.e.,* a soil/topographic index. This index can be applied to calculate storage deficits at each point in the catchment and therefore to calculate saturated contributing areas. This distributed approach is not considered further here. Although Topmodel can simply be viewed as a lumped conceptual model, it is inconsistent with the simplification into sequential domains as used in the model structures applied in this monograph.

The interception process can be important in order to estimate the overall water balance. It is therefore sometimes explicitly modelled, even in simple model structures (e.g., Jakeman *et al.*, 1994). The most common approach to represent interception is by using an overflow storage, which usually additionally empties through evapotranspiration at the potential rate (e.g., Moore, 1999). Gross precipitation enters the storage and the overflow forms the net precipitation which reaches the soil moisture accounting component of the model structure.

Another approach was used by Jakeman *at el.* (1994), who simply subtracted a constant value from the gross precipitation (1–4 mmd^{-1}) for a forested catchment in the USA (Coweeta, North Carolina).

2.4.2 Routing

The soil moisture accounting component produces that part of the rainfall that contributes to runoff, usually called effective or excess rainfall. One or more routing components are typically applied to account for retention and translation effects that occur when the contributing rainfall moves to the catchment outlet via different pathways. Even in complex models, these are often represented by relatively simple structures.

The conceptual element most commonly used to describe this transfer from effective rainfall to runoff is *the conceptual reservoir* or *conceptual store*. The behaviour of this reservoir can be described by combining a storage function and a mass balance equation. The storage function describes the relationship between outflow of the reservoir and the amount of water stored,

$$S(t) = a \cdot Q^n(t) \tag{2.4}$$

where $S(t)$ is the storage [L] at time t, $Q(t)$ is the outflow [LT^{-1}] at time t, a is the storage coefficient [$L^{1-n}T^n$], and n is the coefficient of non-linearity [-].

The main cause of the non-linearity of the relationship between rainfall and runoff is the antecedent moisture condition (Beven, 2000a, p.46). If this can be accounted for in a non-linear SMA component, then it is often possible to approximate the remaining transfer from effective rainfall to streamflow by a linear relationship (e.g., Jakeman and Hornberger, 1993). The reason why this is feasible is that errors in the calculation of effective rainfall are usually much larger than those related to the assumption of linear routing (Beven, 2000a, p.28). This leads to the most common form of the conceptual reservoir, the linear reservoir, *i.e.*, $n = 1$. In this case, parameter a becomes the residence time T [T], and the outflow of the reservoir is directly proportional to the storage. The advantage of the linear reservoir model is its computational efficiency. It can be shown that the linear reservoir is identical to a first-order discrete-

time Transfer Function (TF; Lees, 2000), which can be defined to represent any combination of linear reservoirs connected in parallel and/or series, using partial fraction expansion to perform the decomposition (Young, 1992).

The advantages of the representation of a linear conceptual reservoir in TF form include the availability of powerful system identification techniques for optimal parameter estimation, and increased structural flexibility (Lees, 2000). One system identification technique that has been successfully used to identify TF models in the context of rainfall-runoff modelling is the Simple Refined Instrumental Variable technique (SRIV, e.g., Young, 1985; Jakeman *et al.*, 1990; Young, 1992; Young *et al.*, 1996).

The number of reservoir elements required is, amongst other things, dependent upon the modelling time scale selected. A large number of studies have shown that the most common configuration identified for a daily time scale when using the SRIV technique, is two reservoirs in parallel (e.g., Young, 1992; Jakeman and Hornberger, 1993; Lees, 2000). This structure is commonly used in rainfall-runoff modelling (Moore, 1999), although the use of the TF approach ensures that a parallel structure will not be used unless the data support this level of complexity, and indeed conversely more complex structures are sometimes identified (Lees and Wagener, 2000a, 2000b). In the common situation where a parallel structure is objectively identified, the two reservoirs can be considered to represent a quick and a slow response component. These are often interpreted as quickflow and baseflow processes, although these two components aggregate a number of hydrological pathways as described earlier (Hornberger *et al.*, 1998; Ward and Robinson, 2000). A single reservoir is usually sufficient in the case of a coarser discretization in time (e.g., monthly, Jolley 1995) or when a baseflow component is absent (e.g., in the case of ephemeral rivers [Jakeman *et al.*, 1994]). Physical realism of the selected routing structure is an important criterion if the final aim of the modelling exercise is regional transfer of model parameters.

If the slow response component is mainly associated with contributions from groundwater flow, use of a linear approach assumes that the outflow of the groundwater reservoir is directly proportional to the hy-

draulic head (Hornberger *et al.*, 1998, p.164–165). Cases where this assumption is valid include confined aquifers with constant thickness (Darcy's law), and unconfined aquifers where the variation in flow depth is small, *i.e.*, the impermeable layer is far below the riverbed (Chapman, 1999). However, Wittenberg (1999) found that the behaviour of shallow groundwater reservoirs might be more realistically represented by a non-linear reservoir.

Additional evidence for the need to use non-linear reservoirs is shown by the fact that in some cases three parallel linear reservoirs are required to adequately fit the catchment response (Lees and Wagener, 2000b, 2000c). In these cases, one store normally is used to represent the quick response, while the slow response is divided into two stores with different residence times. One interpretation of this result is that two independent aquifers drain to the river (Wittenberg, 1999). This is often unlikely, and the use of a non-linear reservoir to represent the slow response, in combination with a linear reservoir for the quick response, may be more reasonable.

It has also been observed that UK chalk aquifers show a non-linear transmissivity (Arch, 1997). Fissures are the main cause of high hydraulic conductivity of the chalk. These are generally solution features, most extensively developed near the active zone of water table variation. However, these fissures become more and more closed with depth due in part to increasing compression, leading to a decrease in permeability.

The fact that some catchments show sub-surface losses can also be accounted for through an appropriate routing component. One possibility is a linear reservoir with different outlets (e.g., Chapman, 1999; Sugawara, 1995; Moore, 1999). The outflow from the bottom outlet is the portion of effective rainfall that does not contribute to streamflow but accounts for the subsurface losses from the catchment.

2.5 Local Modelling Procedures

Various modelling procedures for the use of conceptual rainfall-runoff models have been suggested (e.g., Beck, 1981a; Dooge, 1981; Anderson and Burt, 1985). Local procedures vary with the particular circumstances

of each case. However, a number of steps can commonly be found in most modelling procedures when applying a rainfall-runoff model to a particular catchment: these are model structure selection, sensitivity analysis, calibration, validation and prediction (Fig. 2.2).

2.5.1 Model structure selection

A model structure appropriate for the envisaged modelling purpose, the given catchment characteristics and (time-series) data has to be selected (Wagener, 1998). The modelling purpose defines aspects such as which hydrological processes need to be considered (e.g., is a simulation of interflow processes required for water resource assessment at a monthly time-step?) and what modelling time step is required. The available data enable a certain degree of causality of process description to be used and allow a particular minimum temporal resolution. The catchment characteristics (and the hydrologist's perception of the hydrological system) are important criteria to determine what type of process description is suitable. However, subjective criteria such as the availability and experience with a particular modelling code, *i.e.,* a set of mathematical equations implemented in a piece of software, or the cost to purchase or develop a code might be of equal importance in the decision making process.

Woolhiser and Brakensiek (1982, p.15) conclude that objective methods of choosing the best model (structure) have not yet been developed, so this choice remains a part of the art of hydrological modelling. However, future modelling tools might include knowledge-based approaches to support the hydrologist with respect to selecting an appropriate model structure (Wagener, 1998).

2.5.2 Sensitivity analysis

Sensitivity analysis is an investigation of how sensitive the model output is to changes in the parameter values. It can also be used to test the influence of other aspects such as initial or boundary conditions. The analysis can take place at two different stages within the modelling process.

First, it can be placed before the calibration step to identify the most important parameters, and therefore model components, with respect to

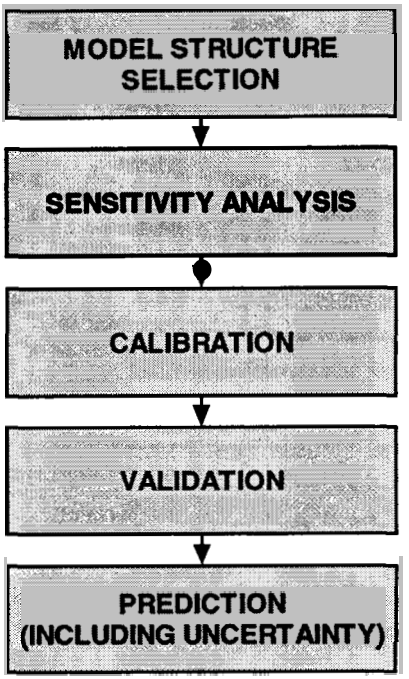


Fig. 2.2 Conventional (simplified) procedure for local rainfall-runoff modelling

the model's performance. Insensitive parameters can then be fixed to a suitable value to decrease the dimensionality of the calibration problem (e.g., Mein and Brown, 1978; Gupta and Sorooshian, 1983; Abdulla *et al.*, 1999; Bastidas *et al.*, 1999). The model performance will often only decrease marginally when this step of complexity reduction is performed sensibly, *i.e.*, when parameter interaction is considered (Bastidas *et al.*, 1999; Wagener *et al.*, 1999; Appendix A). Osidele and Beck (2001) put this approach into a formal framework of hypothesis testing for model structure identification.

The second option is to perform a sensitivity analysis after the calibration step to estimate whether the parameters are identified well or poorly, which is indicated by the response surface slope (Clarke, 1973). The response surface is the topographic surface defined in the $n + 1$-dimensional space by a criterion of model performance (objective function), plotted with respect to the n model parameters. Sorooshian and

Arfi (1982) and Sorooshian and Gupta (1995) describe the use of Taylor series expansion to approximate the response surface around a best parameter set to identify the *region of indifference*, the range in which a parameter can be changed without effect on the model output.

Possible techniques for pre- or post-calibration sensitivity analysis are the perturbation, the factorial, the Fourier Amplitude Sensitivity Test (FAST), or the Regional Sensitivity Analysis (RSA) methods. Bastidas *et al.* (1999) provide an overview of these techniques and examine their usefulness for analysing large-scale hydrological models. The common weakness of all these approaches is that they assume independence between parameters, an assumption rarely justified in RR modelling.

The approaches to sensitivity analysis most often applied to RR models are probably the perturbation and the RSA methods. In the first approach, individual parameters are perturbed over their feasible range using a particular step size, while the remaining parameters are kept fixed at *optimal* values. The effects of these perturbations on the model output, e.g., in the form of changes in a particular OF, can then be measured. An application example of this approach can be found in Rogers *et al.* (1985). This approach however assumes that a single best parameter set can be estimated, an assumption that is questionable for different reasons as explained below. The RSA approach (examined in detail in Chapter 3 as an option within the Monte Carlo Analysis Toolbox, Wagener *et al.*, 2002a) in essence splits parameter sets into one group, which results in model simulations in accordance with observed system behaviour, and a second group for which this is not the case. Differences in the resulting two distributions for individual parameters suggest that the model performance is sensitive to changes in this parameter.

2.5.3 Calibration

As noted earlier, one criterion for a model structure to be classified as conceptual is the requirement that some of its parameters must be estimated through calibration against observed system output. This can be achieved in a single step in cases where the parameters enter the model linearly, *i.e.*,

$$Q = \theta M(I) \tag{2.5}$$

where Q is the streamflow (or any other system output), θ is the parameter vector, M is the model structure (*i.e.,* a collection of functions which can be linear or non-linear), and I is the model input (e.g., precipitation). Solutions can be derived using for example Single Value Decomposition (SVD; Gershenfeld, 1999, p.123).

However, in most RR model structures, the parameters appear inside the constituent functions, *i.e.,* they enter non-linearly,

$$Q = M(\theta, I) \tag{2.6}$$

The consequence is that an iterative solution is required and it is not possible to determine with certainty whether the best solution has been found (Gershenfeld, 1999, p.123).

A large number of calibration procedures to find such a solution in an iterative fashion have been tested over almost four decades (probably starting with Dawdy and O'Donnell, 1965). While varying greatly in detail, they generally consist of the following four components:

- objective function,
- calibration data,
- adjustment strategy,
- termination criterion.

2.5.3.1 Objective function

Visual inspection of calculated and observed hydrographs is a very important tool in any calibration procedure. However, for longer periods of manual calibration and for automatic calibration in general, one must specify an aggregated measure of the model performance, a so-called *objective function* (also sometimes called *loss function*). In RR modelling these are either measures derived from statistics or based on hydrological aspects of the model performance.

Gershenfeld (1999, p.116) describes how statistical performance measures can be derived from a Bayesian basis by asking what is the most likely parameter set θ, given the selected model structure m and state observations d, in other words, the parameter set θ that can be cho-

sen to maximize the probability $p(\theta \mid d,m)$. Using Bayes rule, this task can be written as

$$\max_{\theta} p(\theta \mid d,m) = \max_{\theta} \frac{likelihood \cdot prior}{evidence} \tag{2.7}$$

The three components form a *likelihood* measure, which estimates how well the model predictions and system observations agree with (based on a specific error model), *prior* knowledge on the suitability of particular parameter values within the selected model structure, and the *evidence* that measures how well a particular model structure can reproduce the data (considering all possible models, *i.e.,* parameter sets and model structures).

Assuming (1) that only one model structure is examined, *i.e.,* dropping the conditioning on m; (2) that a prior on the likelihood of the data set is not considered (only required if different data sets of varying degrees of quality are used) and (3) a uniform prior distribution on the model parameters, leads to a *maximum likelihood* (ML) approach (Gershenfeld, 1999, p.117),

$$\max_{\theta} p(\theta \mid d) = \max_{\theta} p(d \mid \theta) \tag{2.8}$$

Assuming further that (1) the residuals, *i.e.,* the differences between observed and calculated system output, are independent and identically distributed (*i.i.d.*), (2) the residual distribution has a homogeneous variance, and (3) the residual distribution follows a normal distribution with mean zero, leads to the simple least squares estimator (SLS; Gershenfeld, 1999, p.118). In RR modelling the SLS most often appears in form of the root mean squared error (RMSE), which has the advantage of being in the same units as the observed system output,

$$RMSE(\theta) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y_t - \hat{y}_t(\theta)\right)^2} \tag{2.9}$$

where $\hat{y}_t(\theta)$ is the calculated flow at time step t using parameter set θ, while y_t is the observed flow at time step t. Variable N is the number of time steps considered for analysis.

Nash and Sutcliffe (1970) derived a normalized form of this measure, the Nash-Sutcliffe efficiency (NSE), which allows, for example, comparison of a model's performance over different catchments,

$$NSE(\theta) = 1 - \frac{\sum_{t=1}^{N} \left(y_t - \hat{y}_t(\theta)\right)^2}{\sum_{t=1}^{N} \left(y_t - \bar{y}\right)^2} \tag{2.10}$$

where $\bar{y}$ is the mean of the observed flows (Legates and McCabe, 1999).

One problem with these two measures is their use of squared residuals which emphasizes the performance during high flow periods (Legates and McCabe, 1999; Wagener *et al.*, 2002a, 2003b). Sorooshian and Dracup (1980) remarked that the errors in the actual flow would be heteroscedastic under the reasonable assumption that the absolute uncertainty in the rating curve will be highest at the high flow end. They suggested transforming the (input and output) data accordingly using a Box-Cox transformation. This transformation can be done explicitly, *i.e.*, transforming the data before calculating the OF, or implicitly, *i.e.*, as an integral part of the OF (see equations for both approaches in Chapter 3). They derived the Heteroscedastic Maximum Likelihood Error (HMLE) measure, which does the latter. The general result of using the HMLE is a more balanced performance over the whole hydrograph and a reduced performance during periods of high flows compared to the RMSE (e.g., Yapo *et al.*, 1996; Gupta *et al.*, 1998).

A wide range of statistical and hydrological OFs is available. However, while different studies have tried to assess the suitability of different performance measures (e.g., Martinec and Rango, 1989; Diskin and Simon, 1977; ASCE, 1993; Gupta *et al.*, 1998), it ultimately remains a subjective decision of the hydrologist to select one or more measures, suitable for the task at hand.

Martinec and Rango (1989) and Atkinson (2001) list examples not purely derived from statistical considerations. They are for example OFs in the form of deviation of runoff volumes or of the aggregated performance over different time-scales (annual or seasonal). Another example is

the use of flow duration curves to judge the quality of the model predictions (see Chapter 5 for example plots).

It has become increasingly apparent that the various OFs emphasize different aspects of the hydrograph and yield different optimal parameter sets. This fact demonstrates one of the main problems in RR modelling, the presence of model structural inadequacies. Despite a different underlying assumption, current model structures are not capable of reproducing all aspects of the hydrograph with a single parameter set (see Yapo *et al.*, 1996; Gupta *et al.*, 1998; Boyle *et al.*, 2000; Wagener *et al.*, 2001). The hydrologist must sacrifice the fit to one aspect of the hydrograph in order to fit another. A significant trade-off is found between the fit to high and to low flows. Figure 2.3 shows a plot of simulation results from Monte Carlo sampling (based on uniform distributions) of the parameter space. The RMSE during high flow periods (F_H) is plotted against the RMSE during low flow periods (F_L), separated by a simple threshold flow value (the mean observed flow), for a typical Conceptual RR (CRR) model (Wagener *et al.*, 2000). A clear trade-off curve between the two criteria is visible, indicating structural inadequacy. Gupta *et al.* (1998) argued from similar observations that the problem of RR model calibration is inherently multi-objective and that the philosophy of calibration has to be rethought. This problem is analysed in detail in Chapter 4.

There are also different ways of combining multiple-objectives into a single value, which then allow the use of a conventional single-objective optimization algorithm. Seibert (2000) combines two OFs, measuring the fit to streamflow and groundwater levels, into a single fuzzy measure. The result is reasonable, but gives sub-optimal performances of the HBV model structure with respect to both measures. Madsen (2000a, 2000b) also combines two OFs (the RMSE for flows above and below a specified threshold). His combination measure uses the Euclidean distance to achieve a trade-off between the two OFs. This allows the estimation of the best parameter set with respect to each OF, *i.e.*, the extremes and all those values that lie along the trade-off front between them, similar to the one found in Fig. 2.3. This approach is however less efficient with respect to the number of required function evaluations than the use of a multi-objective algorithm (Bastidas, 1998; Vrugt *et al.*, 2003b).

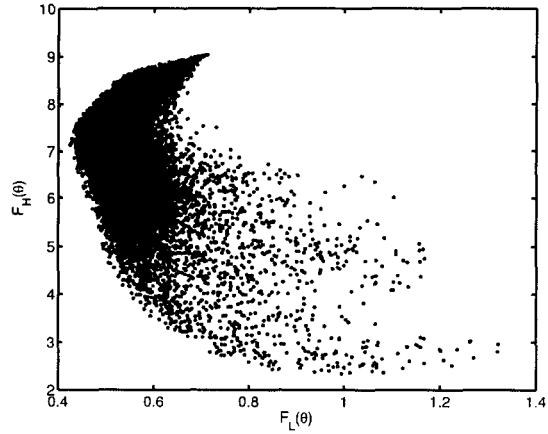


Fig. 2.3 Model population resulting from a Monte Carlo simulation plotted in the 2-dimensional OF space. F_L is the RMSE for the part of the hydrograph classified as low flow; F_H is the RMSE for the high flow part. The optimum value for both functions is in the bottom left corner.

2.5.3.2 Calibration data

The quality of any calibration procedure is very much dependent on the quality and quantity of the time-series data used. The *quality* of the data is a function of the amount of information (with respect to identifying the model parameters) and of the noise present. The quantity required seems to depend on the complexity of the model structure (in terms of number of parameters which need to be estimated) and on the quality and characteristics of the data. Franchini and Pacciani (1991) found that the required length of the calibration data was directly related to the number of parameters to be optimized. Yapo *et al.* (1996) found that 8 years of data are required to make the 13 estimated parameters of the Sacramento model structure independent of the actual period selected, while 3 years of data are suggested to be sufficient for the 6-parameter IHACRES model structure (Jakeman *et al.*, 1993). A shorter time-series (minimum one complete year for continuous modelling) might suffice if it is of high quality and has a high degree of hydrologic variability. Wet years, required to activate all parameters (e.g., thresholds), seem to be of particular importance for the shape of the response surface (Gupta and Sorooshian, 1983; Sorooshian *et al.*, 1983).

2.5.3.3 Adjustment strategy

Initial approaches to find a suitable parameter set used manual, trial-and-error techniques. This technique has evolved over a long time into very sophisticated strategies that commonly use a variety of performance measures and visual comparison of results to guide the search. A very advanced manual procedure is the one used by the National Weather Service (NWS) to calibrate the Sacramento Soil Moisture Accounting (SAC-SMA) model structure (e.g., Brazil, 1988; Boyle *et al.*, 2000). This procedure uses different levels or stages of calibration. Initial parameter ranges are derived from experience with catchments similar in geology, hydrology and climatology. Some of the parameters are then treated as independent entities and adjusted using only particular periods of the hydrograph, e.g., slow recessions to derive baseflow-related parameters. Advanced visualization tools support this process. Complex parameter interactions are then considered in the final and most difficult stage. Relying on visualization, a range of statistical and hydrological criteria, and experience, the hydrologist has to find the most suitable parameter set that achieves a suitable trade-off in the fit between different periods of the hydrograph.

While being able to produce excellent results, the manual approach has a range of serious shortcomings. Boyle *et al.* (2000), for example, report that training for users of the SAC-SMA model structure can require several months and that the calibration of the SAC-SMA to a single catchment may take several hundred hours of work. Manual calibration requires comprehensive understanding of the catchment runoff behaviour and the model structure, and is labour intensive. The termination of the calibration process is based on the subjective decision of the hydrologist and therefore so is the result, and it is difficult to transfer the expertise to another person.

The lack of speed and objectivity in manual calibration led to the introduction of automatic adjustment strategies into RR modelling (Dawdy and O'Donnell, 1965).

Initial approaches used *local schemes*. These start at a single point, *i.e.*, a single parameter set in the feasible parameter space, and follow a

programmed set of rules to find steps of improvement until a specified termination criterion is satisfied. The first algorithms that were applied used only estimates of points on the response surface, *i.e.,* they were direct methods. Popular examples are the *downhill simplex* (Nelder and Mead, 1965; Gershenfeld, 1999, p.157ff.), the *pattern search* (Hooke and Jeeves, 1961), and the *rotating directions method* of Rosenbrock (1960).

Potentially more powerful than direct methods are gradient approaches that also use the first and/or second derivative(s) of the response surface (Bard, 1974, p.117). They obtain more information about the response surface at every iteration step, *i.e.,* the downhill directions at a particular point are known, and should therefore converge faster (Hendrickson *et al.*, 1988). Derivatives can be obtained through numerical approximation (Hendrickson *et al.*, 1988) and even analytical solutions are possible (Gupta and Sorooshian, 1985b). Despite threshold parameters usually present in parametric model structures, Gupta and Sorooshian (1985b) derived analytical derivatives by computing different derivatives for different modes: one mode if the threshold were not exceeded and another if it were. Hendrickson *et al.* (1988) concluded that gradient methods do not perform better than direct approaches in the presence of discontinuities and non-smooth response surfaces typically found in parametric model structures. Moore and Clarke (1981) avoided discontinuities on the response surface caused by thresholds by using a probability distribution of storage elements, instead of individual stores. This allows for the use of a derivative-based optimization approach.

However, as early as 1970, Ibbitt (1970) concluded that the use of local approaches is hindered by the presence of multiple response surface optima. He suggested using random searches to identify promising starting points, a difficult task at that time due to the lack of computer power. His conclusions were subsequently confirmed by other studies (e.g., Johnston and Pilgrim, 1976; Pickup, 1977; Duan *et al.*, 1992).

Approaches to prevent local search procedures from getting stuck at minor (local) optima include the addition of concepts analogous to momentum or temperature to the original algorithm. Adding momentum, for example, to a downhill simplex means adding a fraction of the change of the previous move to the current move (Gershenfeld, 1999, p.161). This can help the simplex avoid minor optima, but also reduces the conver-

gence speed. Simulated annealing (Kirkpatrick *et al.*, 1983; Thyer *et al.*, 1999) adds the idea of temperature, to the vertices of a downhill simplex. The vertices *vibrate* as a function of the current temperature value; this vibration decreases in the course of the optimization, therefore avoiding the problem of getting stuck at local optima in the early stages of the search.

Sorooshian and Gupta (1995) give a summary of reasons for difficulties in automatic calibration with respect to the response surface of parametric structures:

- The parameter space contains several regions of attraction, *i.e.*, several clusters of numerous local optima (see for example Kuczera, 1997).
- The response surface is non-smooth and often shows discontinuities due to the use of thresholds in the model structure. The same is true for the first and second derivative of the response surface.
- The different parameters are not equally sensitive and show a great deal of (often non-linear) interaction.
- The response surface often reveals a non-convex shape around the global optimum.

These findings led to the investigation of so-called global search algorithms in RR modelling (e.g., Brazil, 1988; Wang, 1991; Duan *et al.*, 1992). Global approaches are, among others, pure random sampling methods, adaptive random methods, and evolutionary methods.

Pure random methods are based on sampling a large number of points in the parameter space. The most common approach used in RR modelling, usually as a basis for a detailed uncertainty analysis, is simple random sampling (e.g., Beven and Binley, 1992; Wagener *et al.*, 2003b). In this approach, a (usually) pre-selected number of parameter sets are drawn randomly from prior distributions. Kuczera and Parent (1998) suggest that for continuous multivariate problems only two prior distributions can practically be used. These are uniform or multinormal distributions. However, due to lack of knowledge about the true distribution, practical problems of implementation (Kuczera and Parent, 1998) and the fact that the parameter distributions are often highly skewed, uniform distributions are usually assumed (e.g., Beven and Binley, 1992; Freer *et al.*, 1996; Uhlenbrock *et al.*, 1999). This approach

is quite inefficient and a large number of points have to be sampled and evaluated, especially if high parameter interaction produces a response surface with narrow curving ridges, which can be computationally very demanding in high-dimensional spaces (e.g., Kuczera and Parent, 1998). Schemes have been introduced to ensure that sampling covers the whole search space. Examples are stratified sampling (Press *et al.*, 1992, p.317ff.), Latin Hypercube Sampling (LHS, McKay *et al.*, 1979) or clustering methods (Özdamar and Demirhan, 2000), with the latter requiring a minimum user-specified Euclidean distance between the sampled points.

These methods however do not make use of any information derived by sampling the parameter space. In contrast, *adaptive random methods* are based on the idea of resampling promising areas in greater detail and therefore narrowing down the search space to one or more small areas (e.g., Spear *et al.*, 1994; Schütze *et al.*, 2002). Examples are the Adaptive Random Search (ARS, Masri *et al.*, 1980; Pronzato *et al.*, 1984; used e.g., by Brazil and Krajewski, 1987; Brazil, 1988) and the Controlled Random Search (CRS, Price, 1987; used e.g., by Klepper *et al.*, 1991) methods.

Evolutionary algorithms are probably the most commonly applied global optimization methods in RR modelling, apart from the Uniform Random Search (URS). The main approaches found are Genetic Algorithms (GA, Goldberg, 1989) and the Shuffled Complex Evolution Algorithm (SCE, Duan *et al.*, 1992; 1993; 1994; Sorooshian *et al.*, 1993). These two approaches randomly sample an initial parameter population (usually based on a uniform prior distribution) and then improve its members by different evolution schemes.

The GA uses a number of steps that are related to the theory of evolution to improve a model population. These are usually fitness, reproduction, crossover and mutation steps (e.g., Gershenfeld, 1999, p.164ff.). The steps allow the algorithm to be very flexible, but also make it difficult to find the optimum algorithm parameter settings for a particular problem. A GA is therefore not a single algorithm, but a generic tool. Wang (1991) introduced the use of GAs to RR modelling, applying the method to optimize the Xinanjiang model structure.

The SCE is probably the currently most popular algorithm for the calibration of parametric RR model structures. It combines elements of the earlier-mentioned downhill simplex and CRS, with competitive evolution (Holland, 1975) and the idea of complex shuffling. This algorithm is outlined in Chapter 3. It has been successfully compared to other algorithms in a range of studies, for example to a multi-start simplex (Gan and Biftu, 1996; Kuczera, 1997), a downhill simplex (Gan and Biftu, 1996; Abdulla *et al.*, 1999), to simulated annealing (Abdulla *et al.*, 1999; Thyer *et al.*, 1999); to a multi-start quasi-Newton approach (Kuczera, 1997), and to a GA (Kuczera, 1997). The main reason for its good performance is the use of complexes that share information, thereby ensuring that the search is really global.

These adaptive random search and evolutionary methods are applied to find a global optimum. In contrast to the pure random search approaches, they do not produce an estimate of the resulting posterior parameter distributions, a requirement for an uncertainty analysis. Pure random search approaches on the other hand can be very inefficient and require a very large number of sampled points in higher dimensional spaces (Kuczera and Parent, 1998). The use of a Markov Chain Monte Carlo (MCMC) approach which adapts to the true posterior distribution using a random walk (or an independence) chain (e.g., Kuczera and Parent, 1998; Campbell *et al.*, 1999; Bates and Campbell, 2001; Vrugt *et al.*, 2003a) might be more efficient in those cases. One of these approaches is the Metropolis algorithm (Metropolis *et al.*, 1953) which uses a random walk scheme that always accepts better performing parameter sets, but also accepts worse sets with a particular probability (e.g., Kuczera and Parent, 1998).

Multi-objective calibration procedures (required when using multiple OFs due to multiple output variables and/or structural inadequacies, as mentioned earlier) often utilize population evolution algorithms to estimate the so-called *Pareto set*. A parameter set is Pareto optimal (also termed *noninferior* or *nondominated*) if improving one OF leads to a reduction in performance with respect to at least one of the remaining OFs (Chankong and Haines, 1993).

Both the GA and the SCE can be adjusted to perform in a multi-objective mode in order to estimate a Pareto set. Yapo *et al.* (1998) in-

troduced a multi-objective modification of the SCE algorithm, called MOCOM (Multi-Objective Complex Evolution Method). MOCOM is a general-purpose global optimization algorithm capable of optimizing a model population simultaneously with respect to different OFs in a single optimization run. A detailed description and explanation of the method are given in Yapo *et al.* (1998) and Bastidas (1998); some details can also be found in Chapter 4. Liong *et al.* (2001) and Khu (1998) show how GAs can be applied in a multi-objective fashion to derive a Pareto set for RR models. Vrugt *et al.* (2003b) recently developed a new multi-objective algorithm which also allows for the consideration of parameter uncertainty. The algorithm is called Multi-Objective Shuffled Complex Evolution Metropolis Algorithm (MOSCEM-UA)

It might sometimes be more appropriate to estimate some parameters independently using manual procedures since they are difficult to calibrate in an automatic way. A typical example is the slow response recession constant k(slow) [T^{-1}] usually present in one form or another in structures for continuous modelling at a daily discretization which can be derived from the recession slope, *i.e.*, k(slow) = q_{t+1}/q_t (e.g., Boyle *et al.*, 2000).

2.5.3.4 Termination criteria

Manual calibration is terminated when the hydrologist is satisfied with the model performance, a subjective decision based on personal judgement. More objective criteria are commonly used to terminate automatic calibration procedures. These are (Sorooshian and Gupta, 1995): (a) *function convergence*, which stops the calibration when the improvement in the OF value is below a user selected threshold value over a number of iteration steps; (b) *parameter convergence*, which analyses whether the best parameter estimate remains within a (again user selected) fraction of its feasible range (for adaptive or evolutionary approaches); the search is terminated if all parameters involved remain in this fraction over a number of iterations; and (c) *maximum number of iterations*, a criterion often used for pure random search approaches. It might also be used as a safety mechanism to avoid excessive model runs when the two other convergence criteria are not fulfilled.

Isabel and Villeneuve (1986) emphasize the importance of a strict function convergence criterion. They found numerous sills when plotting the number of simulations against the OF values of the currently best parameter set (see also Pickup, 1977). A threshold value that is too large can therefore lead to premature convergence of the calibration procedure.

Sorooshian *et al.* (1983, 1993) suggest parameter convergence as the most suitable termination criterion. Their research showed that, while OF values often stabilize relatively quickly, the actual parameter values keep changing considerably (Sorooshian *et al.*, 1983), indicating strong parameter interaction.

2.5.4 Validation

A step often included in modelling procedures is validation or verification (e.g., Dooge, 1978; Beck, 1981a, Beven, 2000a, p.4). However, the definitions of what this step actually represents vary considerably (Anderson and Bates, 2001). Very often, validation is taken to mean testing the model on a data set independent from the one used for calibrating the model structure, commonly in the form of a split-sampling test where a data set is divided into two periods. This is however sometimes considered a minimum requirement for the suitability of a model (Klemeš, 1986).

Klemeš (1986) proposed a more thorough, hierarchical scheme of model validation. It contains four stages in which the model's performance to predict streamflow is tested: (1) a simple *split sample* test as described earlier; (2) a *proxy-basin test*, testing the capability of the model to predict flow in a hydrologically similar catchment; (3) a *differential split sampling test*, where the two data periods selected have different hydrological characteristics (e.g., a dry and a wet period); and (4) a *proxy-basin differential split sampling test*, combining tests (2) and (3). The idea is to test the ability of a model to perform a specific task, not to check its hydrological soundness (Klemeš, 1986).

Applications of this type of validation using only streamflow (e.g., Franchini and Pacciani, 1991; Refsgaard and Knudsen, 1996; Mroczkowski *et al.*, 1997; Seibert, 1999a) have shown that it is often not very discriminative since different models or even model structures pass

when using the same type of data. Mroczkowski *et al.* (1997) subsequently added another stage by testing the performance with respect to additional output variables (stream chloride and groundwater). In one case they found that all three variables were required to objectively show that a model failed the validation.

A slightly different approach is the analysis of the residual properties (e.g., Beck, 1981a, 1981b), assuming the hydrological model behaves like a regression model (e.g., Troutman, 1982). The choice of a particular error model, made implicitly when selecting an OF, often allows analysis of the residuals in order to establish whether they conform to the assumptions made. This can be used to show that a particular error model is (statistically) more appropriate (Yapo *et al.*, 1996) or that the model did not extract all available information from the data. The detection of a correlation between residuals and a forcing variable might for example indicate that not all the information in the data is used and another model component could be identified (Beck, 1981a).

The validation approaches described have, however, a scientific basis that is dubious if the idea is to show that a particular model is correct. Oreskes *et al.* (1994; see also Beven 2002) argue that this idea is ill posed and should be replaced by mere model evaluation, leading to the rejection of all those models which are unsuitable. Different hydrologists have adopted the idea of a more Popperian approach (Popper, 2000) in order to derive a more scientific basis for hydrological modelling (e.g., Wagener *et al.*, 2003b). Popper (2000) suggested scrutinizing theories (models) using all available means and rejecting those that fail. However, verification or validation of a theory as the true one is logically not possible. This topic is discussed in more detail in Chapter 4.

2.5.5 Prediction – including uncertainty

Uncertainty is unfortunately an integral part of any hydrological modelling undertaken. It is therefore important to analyse this uncertainty and its effect on the predicted output variable.

The uncertainty in the modelling process has four major components (e.g., Kitanidis and Bras, 1980a, 1980b; Melching, 1995; Høybye, 1996; Bastidas *et al.*, 1999; McIntyre *et al.*, 2002):

- *Data uncertainty, i.e.,* errors introduced by the measurement itself, by the temporal and spatial discretization of measurements (e.g., the use of point measurements to estimate areal average values of a variable) or by data pre-processing;
- *Model structural uncertainty, i.e.,* simplifications and/or inadequacies in the description of real world processes. The unavoidable deficiencies in the model structure often result in the problem that different parameter sets, (*i.e.,* different models) fit one mode of system response at the expense of other response modes that are reproduced less accurately (Gupta *et al.*, 1998);
- *Model specification uncertainty, i.e.,* the inability to converge to a single *best* model using the information provided by the available data. This uncertainty results mainly from data and model structure uncertainties (Melching, 1995) since (a) the calibration procedure propagates data uncertainty into the model parameters leading to similar model performance with erroneous data and parameters as with true data and parameter values (e.g., Melching *et al.*, 1991); and (b) characteristics of the model structure such as thresholds and interacting parameters result in multiple regions of attraction in the model space and in multiple local optima within those regions, making it difficult to identify the globally optimum model (Duan *et al.*, 1994).
- *Uncertainty due to unknown initial conditions, i.e.,* the states of the model (e.g., moisture content of stores) are usually unknown at the beginning of any calibration or simulation period. However, this uncertainty can often be minimized either by calibrating the initial conditions or by using a warming-up period, which allows the internal states to adjust.

Additionally, one has to keep in mind that even if those uncertainties could be removed, there would still be randomness in the natural processes themselves (Melching *et al.*, 1990). This randomness introduces uncertainty that cannot be reduced. It is assumed here that this uncertainty can be associated with data uncertainties if sufficiently long data records are used.

Different approaches have been proposed to propagate the uncertainty present in the different components of the modelling process into the model predictions (e.g., Keesman, 1989; Melching *et al.*, 1990; Beven and Binley, 1992; Melching, 1995; Høybye, 1996; Gupta *et al.*, 1998).

Conventional approaches to estimate prediction uncertainty, such as first-order analysis, estimate the mean and variance of the predicted variable (e.g., streamflow) by propagating the mean and variance of input or parameters through the model (Melching, 1995). Høybye (1998, p.16) notes that this approach gives accurate predictions of the first two moments of the output variable in cases where the non-linearity of the model and/or the uncertainty of the random variables/parameters are small. However, these assumptions, and the assumption of a multinormal distribution around all uncertain variables/parameters involved, are often not suitable for hydrological models, due for example to the presence of threshold parameters (e.g., Parent and Kuczera, 1998).

Various researchers subsequently introduced methods based on Monte Carlo sampling that relax the assumptions inherent in the approaches described above. These examples are more or less based on the RSA methodology described earlier. Examples are the Generalised Likelihood Uncertainty Estimation (GLUE, Beven and Binley, 1992; Freer *et al.*, 1996), the Monte Carlo Set Membership (MCSM, Keesman, 1989; Van Straten and Keesman, 1991) and the Prediction Uncertainty (PU, Klepper *et al.*, 1991) approach. The most popular of these approaches for RR modelling is probably GLUE (e.g., Uhlenbrock *et al.*, 1999, Lamb *et al.*, 1998). The use of GLUE to estimate prediction uncertainty is an option in the Monte Carlo Analysis Toolbox described in Chapter 3 in detail.

However, the methods described above analyse and propagate parameter uncertainty. Yapo *et al.* (1996) concluded from their research "that the factor currently limiting model performance is model (structural) error". It is therefore advisable to explicitly address the uncertainty originating from model structural inadequacies and errors. The nature of model structural error does not allow the estimation of a probabilistic structure (e.g., in the construction of an appropriate OF) to describe it, because the errors are not random in a probabilistic sense (Gupta *et al.*, 1998). Structural uncertainty does, for example, become visible since

different parameter sets are required to enable the model to reproduce different aspects of the hydrograph.

Shamseldin *et al.* (1997) compared two approaches to combine the output of different model structures into a single prediction. The methods investigated were a weighted average (with the normal average as a special case) and an ANN approach. The idea was that the performance of a single model structure might vary with the response mode of the system modelled, and a combination of different model structures might be more reliable. However, the results were mixed and gave no clear indication whether the approach performed generally better than simply using a single model. The main limitation was that all weights were constant in time. An approach with time-varying weights, which increase in periods for which a particular model structure is more suitable, might be more useful.

See Beck (1987), Melching (1995) and McIntyre *et al.* (2002) for a more detailed description of methods for uncertainty analysis and propagation in hydrological modelling.

2.6 Regional Modelling Procedures

Flow predictions are often required for ungauged catchments, which lack or have very limited records of flow measurement. This is a common problem, even in countries with extensive measuring networks such as the UK with over 1400 gauging stations (Sefton and Howarth, 1998). A local approach, requiring parameter estimation through calibration against observed flow, is not applicable in those circumstances and alternatives have to be found.

The main approaches to estimate conceptual model parameters are through *physical reasoning* (e.g., Koren *et al.*, 2000), *statistical analysis* (e.g., Jakeman *et al.*, 1992), or a mixture of both. Physical reasoning in this context means that parameters are derived from catchment properties, either directly or through empirical equations.

Various authors (e.g., Koren *et al.*, 2000; Duan *et al.*, 2001; Atkinson, 2001) have promoted the idea that conceptual model parameters can, more or less, be estimated directly from catchment (mainly soil) proper-

ties. Koren *et al.* (2000) suggest a procedure based on this assumption for the Sacramento model structure used by the National Weather Service in the USA (the paper by Duan *et al.*, 2001, contains the same example with increased detail). The approach by Atkinson (2001) is basically identical with respect to parameter estimation. However, he uses a much simpler model structure. The idea of Koren *et al.* (2000) is to derive good initial estimates for a subsequent calibration procedure in gauged catchments, but also for ungauged catchments and semi-distributed modelling. The approach has some very appealing elements. However, some of the underlying assumptions will have to be corroborated (or rejected) by future studies.

The first assumption is that it is possible to derive some catchment-scale soil properties, such as field capacity (FC) or wilting point (WP). These values are usually derived from a few point samples analysed at the laboratory scale (probably by making the additional assumption that the value for the dominant soil type in the catchment is the correct one). This makes the use of these values for lumped parameter estimation questionable since "there is generally no theory that allows the estimation of the effective values within different parts of a heterogeneous flow domain from a limited number of small scale or laboratory measurements" (Beven *et al.*, 2000).

Secondly, one has to assume that these soil properties are directly equivalent to lumped conceptual model parameters which are effective values since they are estimated from the integrated (streamflow) response of the catchment, including effects of macropores and the unique features of the catchment (e.g., the unique combination of soil types present).

On the positive side, the approach does not assume that the model parameters are independent, which the statistical approach usually and incorrectly does. Again, the actual empirical relationships derived using the physical reasoning approach will have to be validated in a larger study comparing a large number of parameter sets derived from soil properties and from local calibration.

A problem in testing the procedure suggested by Koren *et al.* (2000) and Duan *et al.* (2001) is that the Sacramento model structure is calibrated in their studies using a manual procedure (NWS, 2001) which can

lead to quite different *optimum* parameter sets depending on the subjective judgement of the individual hydrologist. An automatic approach, estimating the Pareto range of the parameters in which the regional parameters should fall, could be a method to circumvent this problem. There is of course the question of what OFs should be used for this.

The results presented by Koren *et al.* (2000) and Duan *et al.* (2001) are encouraging, and suggest that there is some potential in this approach, at least for a relatively complex conceptual structure such as the Sacramento model that describes the vertical hydrological processes in quite some detail.

The possibility that parameters of lumped parsimonious model structures (which usually integrate a range in space and time even more than complex structures) can be easily related to soil properties seems remote. However, assuming that the parameters are at least constants and representative of inherent properties of the natural system if not measurable properties suggests that it could be possible to relate parameters and properties through statistical inference. This is often called *parameter regionalization* (or generalization) and is therefore the approach followed here. The basic idea is to select a local model structure and subsequently calibrate it to a large number of catchments that are in some way similar to the ungauged site for which no streamflow time-series are available. A statistical relationship between the estimated parameter values and (usually) various catchment characteristics is then established (the regional model) and used to predict flow at the new site.

Procedures for parameter regionalization typically contain the following steps (Fig. 2.4): (a) selection of catchments and their characteristics; (b) selection and calibration of the local model structure; (c) selection and calibration of the regional model structure; and (d) prediction of flow at the ungauged site.

2.6.1 Catchment and characteristics selection

The chances of success of the regionalization are likely to be higher if the catchments considered show a high degree of homogeneity with respect to their hydrological behaviour. Initial studies used approaches such as geographical proximity (e.g., NERC, 1975; Manley, 1978; Mos-

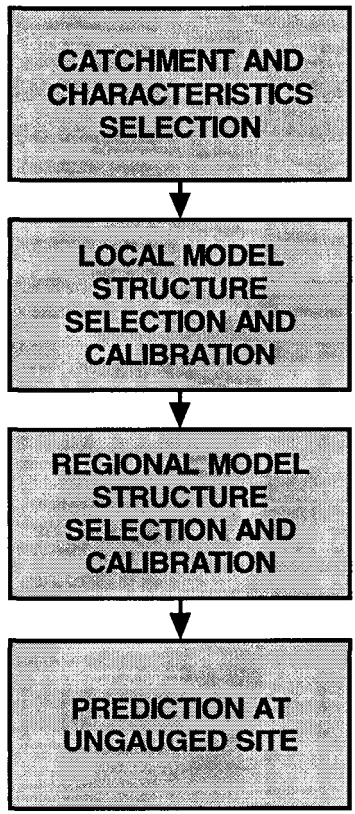


Fig. 2.4 Conventional procedure for regional rainfall-runoff modelling.

ley, 1987; Vandewiele and Elias, 1995) or residual analysis (NERC, 1975; Tasker, 1982) to delineate groups of homogeneous catchments. Residual analysis groups together catchments with regression residuals of similar sign and magnitude. Nathan and McMahon (1990) found that residual analysis did not cluster catchments that were similar with respect to their physical characteristics. Location itself is also an insufficient criterion, since geographically close catchments can be very different, for example with respect to land use or soil type, and catchment characteristics are now commonly included in this type of analysis (e.g., Acreman and Sinclair, 1986; Burn, 1990; Nathan and McMahon, 1990; Dyer *et al.*, 1994; Sefton and Howarth, 1998; Burn and Goel, 2000). These are sometimes combined with standardized flow statistics (e.g., Burn, 1990).

Two main choices must be made in the estimation of homogeneous groups (Nathan and McMahon, 1990): (a) selection of those catchment characteristics (or flow statistics) that define a similar hydrological behaviour, and (b) selection of a linkage algorithm and a distance measure. The most popular approach involves the use of multivariate techniques such as cluster analysis in combination with a Euclidean distance measure (Acreman and Sinclair, 1986; Dyer *et al.*, 1994, p.46; Burn and Goel, 2000). A *region of influence* approach was introduced by Burn (1990; see also Burn and Goel, 2000). He defined a threshold cut-off value for dissimilarity which defines a particular cluster. Hall and Minns (1999) applied an ANN (more specifically a Kohonen network) and fuzzy c-means as an alternative approach.

Nathan and McMahon (1990) provide an overview of the problems related to the derivation of homogeneous groups using multivariate techniques. First, any group of variables (e.g., catchment characteristics) will result in clustering, but with different degrees of relevance. Second, the resulting groups will be a function of the algorithm and distance measures chosen. And third, depending on the approach used, new catchments will usually be allocated to a particular group, regardless of how similar they are.

Visual analysis tools such as Andrew's curves can be used to verify the homogeneity of the resulting groups (Nathan and McMahon, 1990; Dyer *et al.*, 1994) to solve problems two and three. Andrews (1972) developed plots to project *n*-dimensional data into a two-dimensional plane. Dyer *et al.* (1994) used the variables ϕ (catchment characteristics) to produce two-dimensional sin/cos curves. The distance between two curves is proportional to the Euclidean distance metric. Andrew's curves are based on the equation

$$F(i) = \frac{\phi_1}{\sqrt{2}} + \phi_2 \sin(i) + \phi_3 \cos(i) + \phi_4 \sin(2i) + \phi_5 \cos(2i) + \ldots \quad (2.11)$$

where i runs between $-\pi$ and $+\pi$. Similar (n-dimensional) data points will be indicated by similar curves.

The first problem, *i.e.*, the variable selection, is however unavoidable and requires the hydrologist to consider the clustering variables very carefully.

2.6.2 Local model structure selection and calibration

It is commonly assumed that a *general local model structure* exists that is capable of representing all of the catchments included. However, the difficulties in establishing a suitable and unique structure and identifying its parameters have been described in detail above. The lack of identifiability of model parameters has especially been blamed for the lack of success of past regionalization studies (e.g., Moore and Clarke, 1981; Wheater *et al.*, 1993). The reasonable success in regionalizing (very simple) unit hydrograph (UH) model structures for event-based modelling seems to confirm this (e.g., Nash, 1960; NERC, 1975).

Most model structures currently used for regionalization are therefore relatively parsimonious. Examples are the already mentioned IHACRES (Post *et al.*, 1998; Sefton and Howarth, 1998), PDM (Calver *et al.*, 1999; Lamb and Calver, 2002) and abcd (Vogel *et al.*, 1999) model structures. The selected model structure is calibrated to all included catchments and a best parameter set is commonly selected in each case.

2.6.3 Regional model structure selection and calibration

Multiple regression is the most commonly found approach to relate parameter values and catchment characteristics (e.g., Nash, 1960; Herdegeen and Reich, 1974; Jarboe and Haan, 1974; Ando, 1990; Reimer, 1990; Tung *et al.*, 1997; Sefton and Howarth, 1998), sometimes developing different regression equations for different clusters (e.g., NERC, 1975). Stepwise regression is usually applied to develop regional models for the different model parameters. The regression approach is however usually univariate, *i.e.*, only one parameter is considered at a time, and parameter interaction is ignored.

Tung *et al.* (1997) applied multivariate and seemingly unrelated regression analyses to consider the strong correlation between the two parameters of Nash's instantaneous unit hydrograph (NIUH) model. They found that approaches that considered correlation yielded a consistently better performance than univariate regression when regionalizing the NIUH parameters using 42 Taiwanese catchments.

Campbell and Bates (2001) used the Bayesian technique by Campbell *et al.* (1999) to derive local posterior parameter populations, and a subsequent linear regression model to derive a regional relationship to catchment characteristics. The main component is a Markov Chain Monte Carlo approach (see also Kuczera and Parent, 1998) which samples from the posterior local parameter distributions to derive regional parameter estimates. This methodology is applied to regionalize the parameters of the two parameter event-based RORB model structure (Laurenson and Mein, 1995). Campbell and Bates (2001) showed that their approach yields better identified regional parameter estimates than a traditional regionalization approach, which does not consider local parameter interaction. However, they failed to show whether their approach actually provides better predictions of flow events. The possible relaxation of assumptions of linearity of the regional model and normality of distributions within this framework has yet to be explored. The approach also has to be extended to be applicable to more complex model structures. A fully Bayesian approach is beyond the scope of the work presented in this monograph.

A range of regional model structures is discussed and applied in Chapter 5.

2.6.4 Flow prediction at an ungauged site

The conventional approach to flow prediction at an ungauged site is the estimation of a best parameter set using a regional model and deriving a single prediction (e.g., Sefton and Howarth, 1998; Calver *et al.*, 1999). This approach is highly questionable, considering the large uncertainties involved in the local and regional modelling steps. However, the topic of uncertainty in regionalization has been given very little consideration so far. Exceptions are the studies by Lamb and Calver (2002) and Yeh *et al.* (1999).

Lamb and Calver (2002) estimated flood frequency distributions for UK catchments using local and regionalized continuous rainfall-runoff models. They used a uniform random sampling procedure to estimate the uncertainty in the local model. For the regional model, they sampled parameter values from the theoretical (normal) distributions around the re-

gression estimates. The result was that the resulting uncertainties from regional and local estimates were often similar in magnitude.

Yeh *et al.* (1997) estimated uncertainty in regional estimates of the two NIUH parameters, N and K, and the propagation of this uncertainty into the UH using two probabilistic point-estimate methods. Sources of uncertainties considered were the regional (regression) model and its parameters. They found that the uncertainty involved was too high to be ignored.

2.6.5 Variations on conventional regional procedures

Recently, several authors have questioned the conventional approach and attempted to develop improved procedures that address the problem of lack of identifiability of model parameters.

2.6.5.1 Integrated procedure

Vogel *et al.* (1999) compared the conventional regionalization approach, where local and regional model calibration are independent steps, with a regional calibration strategy that optimizes the local model parameters at all sites concurrently with the regional (regression) parameters. This resulted in very high performances of the regional models (coefficients of determination range from 0.90 to 0.99) while the performance for locally calibrated catchments decreased only slightly, but did not yield an improved simulation at ungauged sites compared to the conventional two-step approach. The authors did not give reasons for this result, but it is likely to be related to the fact that different parameter sets often yield identical performances with respect to a selected OF. There might therefore be a certain degree of freedom in estimating a good local parameter set.

Funke *et al.* (1999) proposed a somewhat similar procedure which uses the same effect. They improved the correlation between model parameters and catchment characteristics considerably by adjusting for outliers, using the insensitivity of some parameters with respect to the chosen OF. Two parameters were additionally directly derived from soil

properties based on hydrological process understanding (similar to Koren *et al.*, 2000), reducing the number of free parameters from six to four.

The danger with both approaches is that some model parameters might merely be insensitive to the selected (and possibly inappropriate) OF. Forcing them to take certain values could lead to regional relationships that are artificially made and do not actually exist.

2.6.5.2 Sequential procedure

Lamb *et al.* (2000) introduced a sequential procedure to regionalize a parametric model structure (PDM). The model structure is applied to all catchments using a uniform random sampling procedure. The most identifiable parameter is (subjectively) selected and a regional relationship to catchment characteristics is derived. The parameter is subsequently fixed to its regional value for all catchments and the procedure is repeated till all regional models are established. This reduces the number of parameters by one during every iteration step and therefore increases the identifiability of the model structure. This approach is further analysed in Chapter 5.

2.6.5.3 Indirect procedure

Yu and Yang (2000) used an indirect approach to derive parameter values at ungauged sites without relating them directly to catchment characteristics for Taiwan. Their procedure consists of two steps: (a) the derivation of regional models to estimate ten discharges of different exceedance percentiles of synthetic flow duration curves within a homogeneous region as a function of catchment area; and (b) the calibration of an RR model (HBV) to the flow duration curve. The procedure worked well for the prediction of low flow regimes in their case, but is only likely to be useful in countries with limited gauging networks.

2.7 Summary and Conclusions

Two fundamental questions in conceptual rainfall-runoff modelling have not been successfully answered despite forty years of research efforts:

- What is the appropriate model structure for a given type of hydrological system and a particular modelling task?
- What is the appropriate parameter set within this structure to characterize the unique response features of a particular catchment?

An extensive literature survey (Chapters 1 and 2) has illuminated the problems stated above and highlighted some suggestions for potential ways forward.

Little objective guidance is available for the selection of an appropriate *model structure*. Comparison studies are often based on a single performance measure, include only a limited range of contestants, and their results are difficult to generalize. Despite this, some general observations can be made.

- Simple structures (in terms of number of free parameters) perform as well as complex ones for many purposes.
- The number of parameters that are identifiable (at least with respect to a single OF) is between three and five.
- Many model structures have been developed, but only a limited number of components are used within them.

Local modelling procedures were initially based on the assumption that hydrological models could be treated in a way similar to regression models, *i.e.*, simple optimization with respect to a single OF was considered sufficient. Recent research has questioned the usefulness of this paradigm and some researchers call for more *conceptual* modelling approaches that better recognize the nature of the model structures used. The lack of parameter identifiability has led some researchers to the conclusion that the idea of an optimum parameter set should be abandoned in favour of a population of acceptable (behavioural) parameter sets or models. This leads to uncertainty which has to be considered. Additionally, the influence of model structural error appears to be higher than has been recognized so far. However, no suitable approach to assess this uncertainty explicitly is currently available.

Some applications of *regional procedures* for conceptual model parameter regionalization can be found in the literature, but few analyse the elements (and possible alternatives) of those procedures in detail. Nor

have the presence and influence of uncertainty been appropriately considered in regionalization.

The following three-step approach has been chosen to address the areas reviewed above in the following chapters:

- Develop an RR modelling and analysis toolkit to implement and evaluate different model structures.
- Review and improve existing procedures for local modelling, *i.e.,* for modelling gauged catchments.
- Review and evaluate the idea of parameter regionalization, *i.e.,* regional or ungauged catchment modelling.

Chapter 3

A Toolkit for Rainfall-Runoff Modelling

Test everything. Hold on to the good. Avoid every kind of evil.

1 Thessalonians 5, 21-22

3.1 Introduction

As noted in Chapter 1, RR models are well-established tools that are widely utilized in engineering practice. The majority of model structures currently used can be classified as conceptual when the definition of Wheater *et al.* (1993) is applied (see Chapter 1 Introduction for details).

It was concluded in Chapter 2 that conceptual model structures suffer from a number of problems despite their frequent use and development over several decades. A major constraint is the lack of identifiability: different combinations of parameters (e.g., Johnston and Pilgrim, 1976; Beven and Binley, 1992), and sometimes even different model structures (e.g., Uhlenbrock *et al.*, 1999) yield similar results in terms of a defined performance measure or objective function (OF). This results in difficulties in interpreting past behaviour of the catchment system, and hence, difficulties in the propagation of uncertainty into future predictions in the form of wide confidence limits, *i.e.,* a wide range of possible system behaviours (Wheater *et al.*, 1986; Mroczkowski *et al.*, 1997).

The need for model calibration is a major limitation when ungauged catchments have to be modelled. One possible approach to dealing with this problem is to regionalize the parameters of a certain model structure (e.g., Jakeman *et al.*, 1992). Uncertainty in the model parameters or structure (due to a lack of identifiability) significantly limits the use of

models for this kind of regionalization because it is difficult to establish sensible statistical relationships (e.g., Moore and Clarke, 1981; Kuczera, 1983; Wheater *et al.*, 1993). A model structure with identifiable parameters (*i.e.,* with a high *regionalization potential)* is therefore a prerequisite for successful regionalization.

Possible ways to produce better identified models are: (1) reducing model complexity to contain only those components, and therefore parameters, that can be identified from the available data (*i.e., parsimonious modelling*, e.g., Jakeman and Hornberger, 1993; Young *et al.*, 1996); (2) improving use of available information, e.g., using different data periods to identify different parameters or groups of parameters (e.g., Wheater *et al.*, 1986; Dunne, 1999; Wagener *et al.*, 1999; 2001); and (3) using additional information, *i.e.,* multi-response data such as water quality data, groundwater levels, or tracer measurements (e.g., de Grosbois *et al.*, 1988; Kuczera and Mroczkowski, 1998; Seibert, 1999a). It should be noted that using additional output variables is unlikely to be particularly useful in regionalization studies, since multi-response data are not commonly available. This approach is therefore not investigated further here; instead we focus on methods of reducing model complexity and increasing the information that can be retrieved from streamflow measurements.

The RRMT is a toolkit that enables the development, analysis and comparison of model structures of different levels of complexity using all available information. It can be used to facilitate the identification of the appropriate level of complexity that yields a sufficiently high level of performance, while retaining an acceptable level of parameter uncertainty. The toolkit is described in this chapter and a limited modelling exercise is presented to illustrate its use. The toolkit is freely available for non-commercial use and can be downloaded as outlined in Appendices B and C.

3.2 Rainfall-Runoff Modelling Toolbox

3.2.1 General

As suggested above, hydrologists seek the development of a model structure that balances appropriate complexity with model performance and associated uncertainty. The philosophy is based on the recognition that no model structure is suitable for all modelling tasks, and that the appropriate model structure is a function of: (1) the modelling objectives (e.g., required spatial and temporal discretization, relevant response modes to be simulated), (2) the characteristics of the hydrological system under investigation (e.g., dominant processes, response times of the system), and (3) the available data (e.g., possible spatial and temporal discretization) (Wagener, 1998).

A number of *modelling shells* with different levels of complexity can be found in the literature (Overland and Kleeberg, 1993; Woods and Ibbitt, 1993; Leavesley *et al.*, 2002). These systems allow a user to test the suitability of different model components and to combine them in a modular fashion. Components can typically be modified or added if none of the available components fulfils the problem-specific requirements. The Rainfall-Runoff Modelling Toolbox (RRMT; Wagener *et al.*, 1999; 2002a) has been developed, in particular, to facilitate identification of parsimonious, lumped model structures with a high level of parameter identifiability.

The RRMT is a generic modelling framework or shell that facilitates the implementation of different model structures. It can, therefore, be considered to represent a modelling concept rather than a specific model structure. The RRMT is implemented within the MATLAB (Mathworks, 1996) programming environment.

The model structures that can currently be implemented using RRMT are spatially lumped with low or medium levels of complexity (in terms of the number of parameters). Such models can be classified as parametric or hybrid metric-parametric in type (Wheater *et al.*, 1993). The latter type is related to a systems approach to hydrologic modelling (see examples in Jakeman *et al.*, 1990; Jakeman and Hornberger, 1993; Sefton and

Howarth, 1998; Kokkonen and Jakeman, 2001). The aim of this hybrid approach is to use observations (the metric paradigm) and other prior knowledge to test hypotheses about the structure of component hydrological stores (the parametric paradigm) at the catchment scale (Wheater *et al.*, 1993).

The lumped approach is often assumed to be only suitable for small catchments that are relatively homogeneous in terms of soil, vegetation and geology (Blackie and Eeles, 1985). However, experience with lumped models has shown that they can actually provide acceptable performance for a wide range of catchment sizes, *i.e.*, the aggregated response of a catchment can often be effectively characterized by a spatially aggregated model (e.g., Littlewood and Jakeman, 1992; Jakeman and Hornberger, 1993; Jolley, 1997; Perrin, 2000). Blackie and Eeles (1985) suggest that the most important criteria for the suitability of a lumped approach is not the homogeneity of the catchment characteristics, but the stability of the catchment system with respect to the spatial distribution of precipitation, vegetation and soil types.

3.2.2 System architecture

The system architecture of the RRMT is based on a modular structure. The modelling component consists of a moisture accounting module and a routing module (Fig. 3.1); other available modules include optimization, visual analysis, and off-line data processing options. Different approaches can be used to represent each module, and a set of alternatives that is broadly representative of the range of current modelling practice is provided. However, each module (e.g., routing) has a specified input-output structure, and can therefore easily be replaced by a new or modified module, as long as it conforms to the prescribed structure.

The function of the moisture accounting module is to transform rainfall into effective rainfall, and includes a number of different representations of evapotranspiration and soil moisture storage. The routing module simulates the lateral flow processes through various pathways, *i.e.*, overland flow, throughflow and groundwater flow (Ward and Robinson, 2000). An underlying simplifying assumption made here is that the rainfall-runoff relationship can be represented by a sequential architecture of

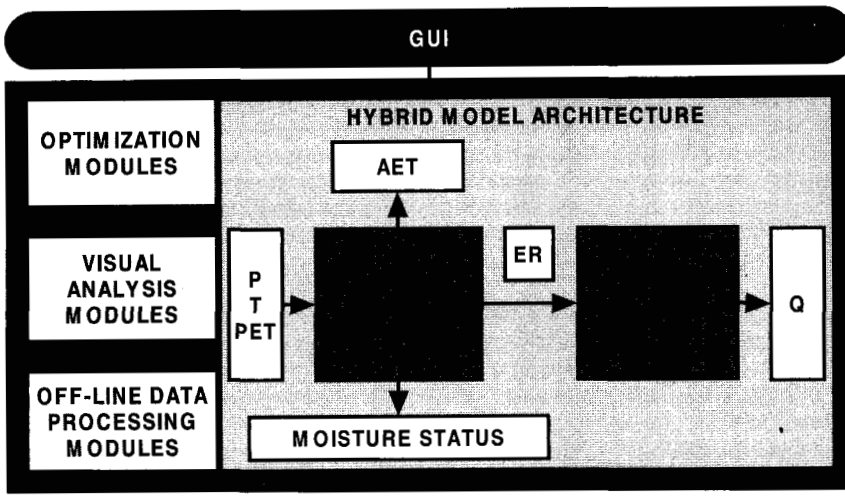


Fig. 3.1 System architecture of the Rainfall-Runoff Modelling Toolbox.

soil moisture accounting and routing modules. While variations on this approach can be found, this assumption is in line with current modelling practice (e.g., Jakeman and Hornberger, 1993; Moore, 1999). A second simplifying assumption is that the effective rainfall can be split into fractions that contribute to different flowpaths and that these fractions are constant in time, rather than being dependent on (for example) catchment wetness or other factors. The usefulness of this assumption is discussed in Section 4.4.2 and recommendations for future alternative options are made in Chapter 7.

A Graphical User Interface (GUI) for the RRMT allows easy access to the toolbox functions (Fig. 3.2). However, since modelling of a large number of catchments may be required (e.g., in regionalization studies), it is also possible to run the model from the MATLAB command line. The user can write a batch file to load data, change model settings, run a calibration, and store the results for many different cases. Within the batch file, all components of the modelling procedure can be varied: the user can select different routing modules or different OFs for the same data, or perform simulations for different years of a data time series.

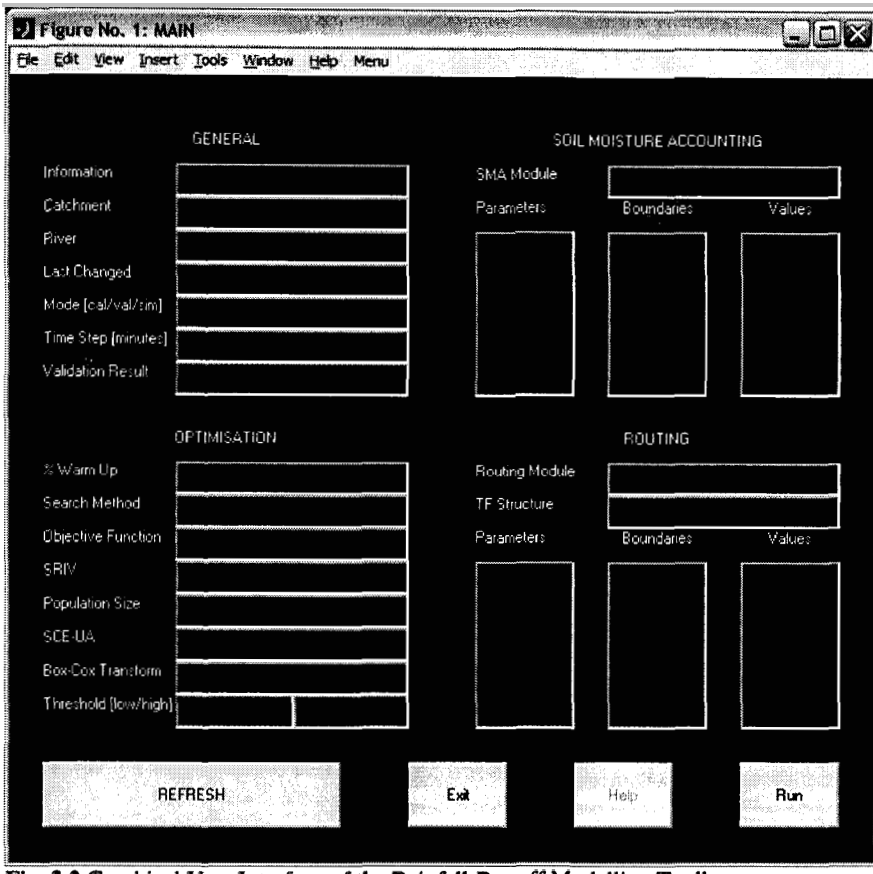


Fig. 3.2 Graphical User Interface of the Rainfall-Runoff Modelling Toolbox.

3.2.3 Soil moisture accounting modules

The soil moisture accounting (SMA) module partitions the incoming rainfall into losses (through evapotranspiration and associated storage) and output from the catchment system (*i.e.*, effective rainfall). An implicit form of the general water balance equation describing this process is (e.g., Hornberger *et al.*, 1998; Blackie and Eeles, 1985)

$$\frac{dV}{dt} = P - Q_{total} - AET \tag{3.1}$$

where V [L^3] is the volume of water stored in the catchment, P [L^3T^{-1}] is the precipitation rate, Q_{total} [L^3T^{-1}] is the rate of surface and subsurface runoff, and AET [L^3T^{-1}] is the rate of actual evapotranspiration. Precipitation combines moisture system input in various forms, e.g. rainfall or snow. However, because none of the catchments used in this study required the explicit modelling of snow, only the term rainfall is used in the following discussion.

The moisture accounting module allows the user to develop and test models having varying levels of complexity. Approaches currently implemented in the RRMT (Fig. 3.3) range from parametric water balance structures such as those based on Penman's drying curve concept (Penman, 1949; for example, see Jolley, 1995) or the catchment moisture deficit (cmd; Evans and Jakeman, 1998), to metric loss functions such as the catchment wetness index (cwi, Jakeman *et al.*, 1990; Jakeman and Hornberger, 1993). These approaches are typical of parsimonious moisture accounting components found in many popular model structures.

The simple cwi loss function is related to the well-known Antecedent Precipitation Index (API, Fig. 3.2a, b). The proportion of rainfall r_k contributing to runoff (*i.e.*, the effective rainfall u_k) at every time-step k, is determined by the cwi, ms, which is calculated as the mean of the indexes for time-steps k and k-1 (an indication of the soil moisture state of the catchment, ranging from zero to one).

$$u_k = \frac{1}{2}\left(ms_k + ms_{k-1}\right) \cdot r_k \tag{3.2}$$

The index ms_k is calculated using the equation

$$ms_k = v \cdot r_k + \left[1 - \frac{1}{\tau(t_k)}\right] ms_{k-1} \tag{3.3}$$

where v is a factor that is adjusted to ensure that the total volume of modelled effective rainfall equals the total volume of observed streamflow. The parameter v is not calibrated but calculated, but is calculated explicitly from the data. Writing this equation in a slightly different form provides more insight into the model structure,

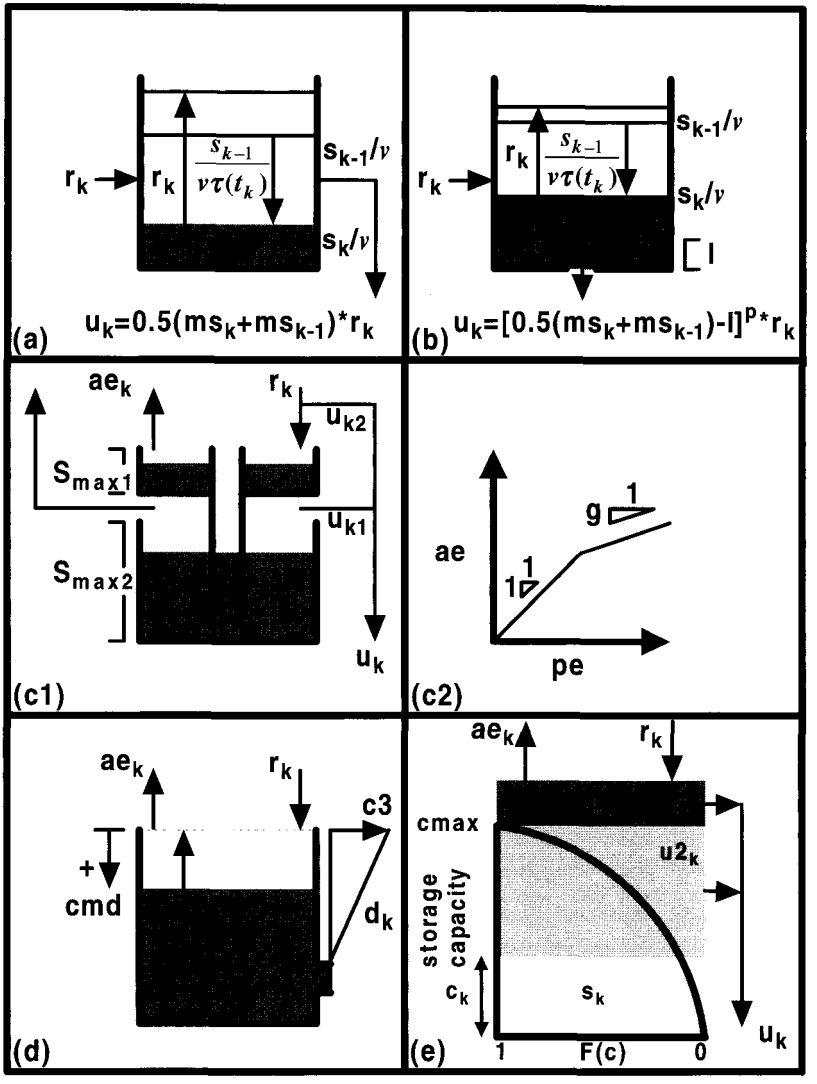


Fig. 3.3 Schematic plots of the different moisture accounting modules available: (a) catchment wetness index (cwi), (b) modified cwi, (c1) conceptual Penman model structure, (c2) Penman drying curve, (d) catchment moisture deficit model structure (cmd), (e) storage capacity distribution function.

$$\frac{ms_k}{v} = r_k + \frac{ms_{k-1}}{v} - \frac{ms_{k-1}}{v\tau(t_k)} \tag{3.4}$$

Equation 3.4 shows that the *modulated system moisture state* at time-step k equals the sum of the rainfall input at time-step k and the modulated system state at time-step k-1, minus the depletion by losses to stream and evapotranspiration described by $\tau(t_k)$ (see Fig. 3.3a).

The depletion is related to temperature (used as a surrogate for potential evapotranspiration), by the following equation,

$$\tau(t_k) = \tau_w \cdot \exp[(refp - t_k)mf] \tag{3.5}$$

where the reference termperature *refp* is usually fixed to a nominal value depending on the climate (e.g., *refp* = 10°C for England and Wales [Sefton and Howarth, 1999], *refp* = 20°C for warmer climates [Jakeman *et al.*, 1994]). The equation contains two parameters that have to be calibrated: τ_w, the time constant of catchment losses at *refp*, and the temperature modulation factor *mf*, which relates a unit change in temperature to the change in loss rate. Note that the temperature can be replaced by potential evapotranspiration when available (e.g., Niadas, 1999). Also, substantial performance improvement can sometimes be found when *refp* is calibrated rather then fixed.

Modifications of the cwi have been developed to account for specific hydrologic catchment characteristics. For example, a power transformation p to ms_k has been used successfully in catchments with a very flashy response, and a threshold l for ms_k has been applied, below which no effective rainfall is produced, to enable the modelling of ephemeral rivers (Fig. 3.3b; Jakeman *et al.* 1994; Ye *et al.*, 1997, 1998).

The basic element of many lumped model structures is a single storage element or bucket (Fig. 3.4; Manabe, 1960), which describes the local water balance as follows:

$$u_k = \max[r_k - (c\max - s_{k-1}) - ae_k, 0] \tag{3.6}$$

where r [L] is the rainfall, $c\max$ [L] is the size of the storage element or the storage capacity, s_{k-1} [L] is the initial moisture content, ae [L] is the

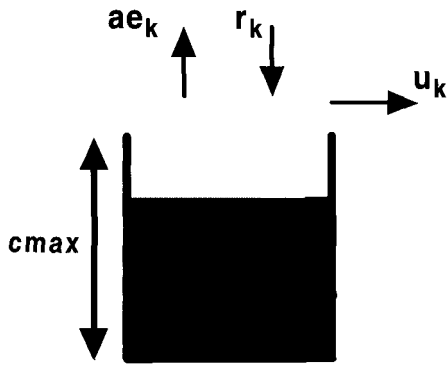


Fig. 3.4 Single storage element defined by the size parameter cmax. The fluxes are actual evapotranspiration ae_k, rainfall r_k and effective rainfall u_k. All at time step *k*.

actual evapotranspiration and *u* [L] is the effective rainfall, while *k* is an index of the time-step.

The catchment moisture deficit (cmd, Fig. 3.4d) approach used by Evans and Jakeman (1998; see also Kokkonen and Jakeman, 2001) is based on a modified version of this simple bucket, described by the following equation of the catchment water balance,

$$cmd_k = cmd_{k-1} - r_k + ae_k + d_k \tag{3.7}$$

where *cmd* is the catchment moisture deficit, *r* is the precipitation, *ae* is the (actual) evapotranspiration loss, *d* is the drainage and *k* is the time-step.

The drainage *d* is introduced due to the assumption that there is a certain amount of runoff to the stream, even when a positive catchment moisture deficit exists (Evans and Jakeman, 1998). This can, for example, take into account the portion of rainfall falling directly on open water, or drainage from the soil near to the stream, which can occur even in periods without rainfall. This phenomenon has already been recognized by others, as for example stated by Hough *et al*: (1997, p.23): "Many experiments have shown that drainage continues into early summer in some soils even when the soil moisture deficit has become substantial near the surface."

The drainage can be calculated as follows:

$$d_k = \begin{cases} -\frac{c_3}{c_4} cmd_k + c_3 & for \quad cmd_k < c_4 \\ 0 & for \quad cmd_k \geq c_4 \end{cases} \tag{3.8}$$

where c_3 is a parameter equal to the maximum drainage that can occur while a moisture deficit exists and c_4 is the maximum cmd that can occur before water ceases draining to the stream.

The equation for (actual) evapotranspiration is based on the assumption that the effects of vegetation can be represented by taking the *surface resistance* into account, which itself can be estimated by creating a relationship with the available soil moisture (Evans and Jakeman, 1998),

$$ae_k = c_1 t_k \exp(-c_2 cmd_k) \tag{3.9}$$

where t is the temperature (can be replaced by pe if available), and c_1 and c_2 are scaling parameters.

To make the calculation procedure easier, a simplification as used by Evans and Jakeman (1998) is also applied here. The term cmd_k is replaced by $[cmd_{k-1} - r_k]$ in the equations to calculate evapotranspiration and drainage.

The effective rainfall can then be calculated as follows,

$$u_k = \begin{cases} d_k & for \quad cmd_k \geq 0 \\ d_k - cmd_k & for \quad cmd_k < 0 \end{cases} \tag{3.10}$$

The second part of Equation 3.10 ($u_k = d_k - cmd_k$) can also be written as the following, if Equation 3.7 is considered,

$$u_k = d_k - cmd_{k-1} + r_k - ae_k - d_k \tag{3.11}$$

$$u_{kr} = r_k - cmd_{k-1} - ae_k \tag{3.12}$$

Rainfall (minus the evapotranspiration losses) is therefore used to fill the soil moisture store and account for evapotranspiration. Surplus rainfall is assumed to be effective rainfall.

After each time step, the value for cmd_k is checked, and is reset to zero if negative. If the rainfall in a time-step is larger than the existing catchment moisture deficit, then the surplus rainfall *goes to the stream* and the saturated soil remains. Evaporation and drainage will then cause the catchment moisture deficit to begin to increase in the next time-step.

One conceptual water balance component implemented in the RRMT which uses the idea of a near surface and a deep layer is a version of the Penman model structure used by Jolley (1995), based on the Penman drying curve concept (Penman, 1949). The approach assumes that actual evapotranspiration (ae) occurs at the potential (pe) rate wherever water is available in the root zone or root-reservoir (Fig. 3.4c1, c2). The root zone is extended by an additional 25mm to account for the effects of capillary rise. The actual rate decreases to a percentage of the potential rate when this soil zone is depleted. The parameterization of this is a conceptual structure with two stores. The size of the upper store is equal to the *root constant* plus the aforementioned 25mm, and evapotranspiration from it takes place at the potential rate as long as it contains water. Values for the root constant can be selected as a function of vegetation from tables (e.g., Grindley, 1970). However, treating the root constant as a free parameter to be identified by calibration can lead to an improvement in model performance (Sheratt, 1985). The upper store is connected to a lower store of *notional infinite depth* (Moore, 1999) via an overflow mechanism. Actual evapotranspiration continues at a fraction g = 8% ($\approx 1/12$) of the potential rate (as suggested by Penman, 1949) from the lower store, after the upper store is depleted.

The amount of actual evapotranspiration from upper (ae_u) and lower (ae_l) stores is therefore calculated as follows,

$$ae_u = \begin{cases} pe & ,md_1 < s_{\max 1} \\ 0 & ,md_1 = s_{\max 1} \end{cases} \tag{3.13}$$

$$ae_l = \begin{cases} g(pe - r - s_{\max 1} + md_1) & ,md_1 = s_{\max 1} \\ 0 & ,md_1 < s_{\max 1} \end{cases} \tag{3.14}$$

where pe is the potential evapotranspiration, r is the rainfall input, md_1 is the moisture deficit in the upper store, and s_{max1} is the size of the upper store, *i.e.,* the root constant plus 25mm.

The effective rainfall u is produced through two mechanisms: (1) a fraction ρ (usually 15% according to Moore, 1999) bypasses the soil water store to represent effects such as rapid groundwater recharge or

quick runoff response during rainfall, e.g., due to macropores or infiltration excess overland flow (Jolley, 1995),

$$u1 = \begin{cases} \rho(r - pe) & r > pe \\ 0 & r \leq pe \end{cases} \tag{3.15}$$

and (2) saturation excess runoff is produced when both stores are full,

$$u2 = \max\left[(1-\rho)(r - pe) - md_1 - md_2; 0\right] \tag{3.16}$$

where md_2 is the soil moisture deficit of the lower store. Hence the total effective rainfall u can be calculated as

$$u = u1 + u2 \tag{3.17}$$

The main advantage of the Penman model structure is its parsimony.

The predominantly vertical processes described by the presented moisture accounting components outlined above imply spatial homogeneity of the hydrological response. This assumption is usually not valid for entire catchments (or often even individual hillslopes), although good results have been reported in many cases (e.g., Blackie and Eeles, 1985; Jakeman and Hornberger, 1993). A catchment is usually heterogeneous with respect to geology, soil type, topography and vegetation, and therefore with respect to hydrological response. Areas in the close vicinity of rivers, for example, might be near saturation most of the time while elsewhere topography and soil type may hardly allow saturation at all, even after prolonged rainfall.

During rainfall events, increasing portions of the catchment will become saturated and contribute to the generation of quick runoff. A simple approach to account for such spatial variations is a distribution of functional hydrological responses, *i.e.,* a mathematical function describing the distribution of storage elements of different sizes. Of course, this is still a lumped approach since it does not allow for the mapping of those responses back into the catchment (Beven, 2000a). This distribution function approach is the basis of the Probability Distributed Model (PDM, Moore and Clarke, 1981; Moore, 1985; Moore, 1999), ARNO (Dümenil and Todini, 1992), XinanJiang (Zhao and Liu, 1995), and the variable infiltration capacity (VIC, Wood *et al.*, 1992) model structures.

As an illustration, the PDM structure is described in detail here. Moore and Clarke (1981) developed the PDM model structure with the following objectives in mind:

- it contains a small number of parameters that allow physical interpretation,
- its simplicity results in identifiable parameters,
- it provides an explicit form for the calculated streamflow given by the model (*i.e.,* it does not contain thresholds), and
- it supports the application of the body of statistical theory available for parameter estimation, hypothesis testing, confidence interval calculation, and for incorporating prior or additional information.

For a probability distribution of storage elements, the total volume of effective rainfall produced in the catchment is the integral response of all storage elements (Dümenil and Todini, 1992) that can be derived from the equation for a single element (Fig. 3.3) as follows,

$$u = \frac{1}{A}\int_0^{A_c} u\, dA_c = \frac{1}{A}\int_0^{A_c} \left[r - (c\max - c) - ae\right] dA_c \qquad (3.18)$$

where $A[L^2]$ is the total catchment area and $[L^2]$ is the contributing area. The ratio A_c/A therefore describes the fraction of the catchment contributing to runoff production, $F(c)$.

The PDM assumes that the storage elements are distributed according to the probability density function $f(c)$ that is related to a distribution function of storage capacity by

$$f(c) = \frac{dF(c)}{dc} \qquad (3.19)$$

Moore (1999, p.154) reported that the most commonly used distribution of storage capacity for the PDM model structure in the UK is the Pareto distribution (e.g., Kottegoda and Rosso, 1998, p.456). The distribution function of store capacities for this case becomes

$$F(c) = 1 - \left(1 - \frac{c}{c\max}\right)^b, \quad 0 \le c \le c\max \qquad (3.20)$$

and the probability density function can be written as

$$f(c) = \frac{dF(c)}{dc} = \frac{b}{c\max}\left(1 - \frac{c}{c\max}\right)^{b-1}, \quad 0 \le c \le c\max \tag{3.21}$$

where the parameter b describes the degree of spatial variability of storage capacity and *cmax* is the capacity of the largest store. A value of b = 1 means that the capacities are uniformly distributed, while a value of b = 0 leads to the assumption of a constant capacity throughout the catchment (Fig. 3.5).

The Pareto distribution function can be implemented as follows. The maximum combined content of all stores s_{max} can be calculated from the two free parameters, *cmax* and b,

$$s_{\max} = \frac{c\max}{(b+1)} \tag{3.22}$$

A spatially uniform rainfall input filling all the storage elements produces the following effective rainfall,

$$u1_k = \max\left[r_k - c\max + c_{k-1}, 0\right] \tag{3.23}$$

[Note: the rainfall input is here assumed to be uniformly distributed over the catchment. For details on how to apply a spatially distributed input, please see Jolley (1995)].

The remaining net rainfall r^* can be calculated as (Fig. 3.3e)

$$r_k^* = r_k - u_{1k} \tag{3.24}$$

and the new actual capacity is therefore

$$c_k = c_{k-1} + r_k^* \tag{3.25}$$

with the storage content being

$$s_k = s_{\max}\left[1 - \left(1 - \frac{c_k}{c\max}\right)^{(b+1)}\right] \tag{3.26}$$

The additional effective rainfall produced by the stores that are smaller than *cmax* is

$$u2_k = \max\left[r_k^* - (s_k - s_{k-1}), 0\right] \tag{3.27}$$

Finally, s_k is updated by subtracting losses due to evapotranspiration

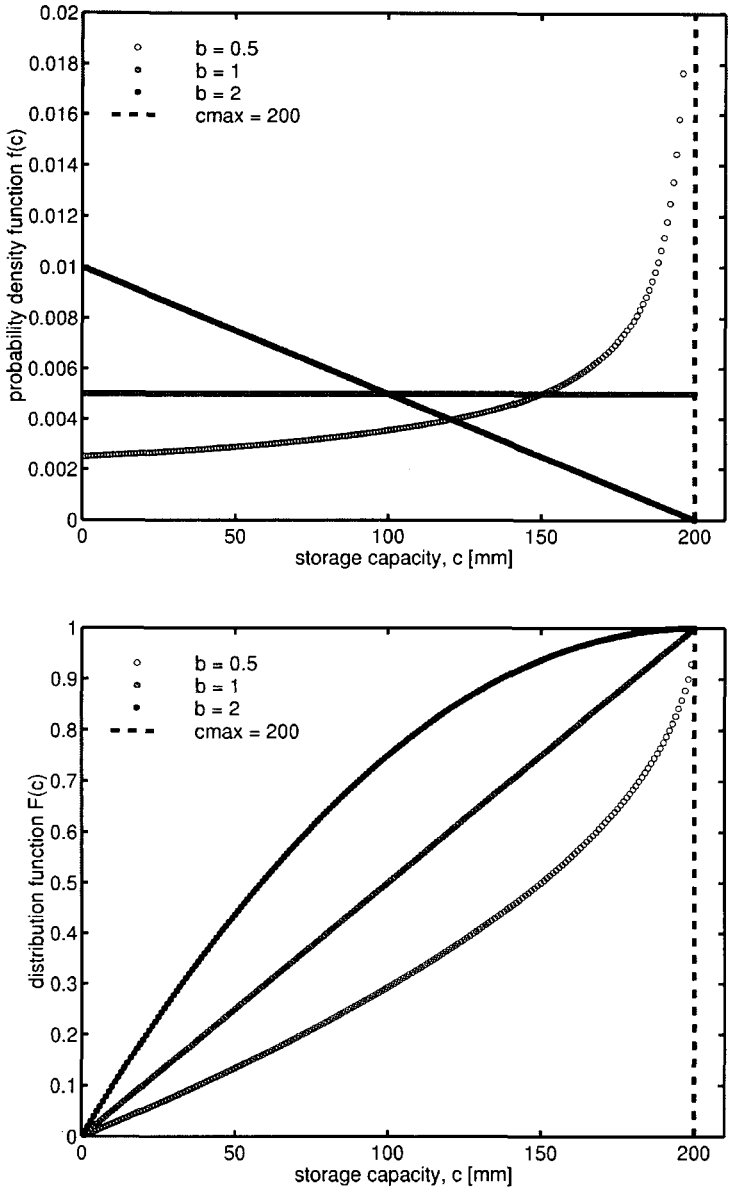


Fig. 3.5 Pareto distribution of storage capacity. The probability density function is shown at the top, the distribution function at the bottom.

$$s_k = s_k - ae_k \tag{3.28}$$

which leads to the new capacity

$$c_k = c \max\left[1 - \left(1 - \frac{s_k}{s_{\max}}\right)^{1/(b+1)}\right] \tag{3.29}$$

RRMT currently provides two different options for the calculation of *ae*. One sets the actual rate equal to the potential as long as sufficient soil moisture is available (as used in Wagener *et al.*, 2001), while the other uses a linear relationship between evapotranspiration and actual soil moisture deficit. These two approaches can be described mathematically as follows:

$$ae_k = \min\left[s_k, pe_k\right] \tag{3.30}$$

and

$$ae_k = \left[1 - \left(\frac{s_{\max} - s_k}{s_{\max}}\right)^{b_e}\right] \cdot pe_k \tag{3.31}$$

The latter relationship is linear for the case $b_e = 1$. However a quadratic form is also sometimes assumed, *i.e.*, $b_e = 2$ (Moore, 1999, p.153). For the case of $b_e = 1$ the equation (3.31) becomes

$$ae_k = \frac{s_k}{s_{\max}} \cdot pe_k \tag{3.32}$$

Note that this formulation assumes that all effective rainfall is produced through overflow of the storage elements. A simple linear relationship is then used to split this effective rainfall into a quick and a slow component, *i.e.*, a constant segmentation. A modified formulation can also be implemented in which recharge to groundwater is computed as a linear function of the soil moisture content (e.g., Moore, 1999; Lamb, 2000)

$$\gamma_k = k_b \cdot s_k \tag{3.34}$$

where k_b is a groundwater recession constant [T^{-1}].

Further increases in the level of model complexity further can allow for the mapping of contributing areas back into the catchment, such as presented in the Topmodel concept and outlined briefly in Chapter 2 (Beven and Kirkby, 1979; Beven *et al.*, 1995).

3.2.4 Routing modules

The moisture accounting component produces that part of the rainfall that is contributing to runoff, usually called effective or excess rainfall. One or more routing components are typically applied to simulate retention and translation processes occurring when the contributing rainfall moves to the catchment outlet via different pathways. Even in complex models, these are often represented by relatively simple structures.

The conceptual (or parametric) element most commonly used to describe this transfer from effective rainfall to runoff is the so-called conceptual reservoir or conceptual store. The behaviour of this reservoir can be described by combining two equations. A storage function describes the relationship between outflow of the reservoir and the amount of water stored,

$$s(t) = a \cdot q^{nl}(t) \tag{3.35}$$

where $s(t)$ is the storage [L] at time t, $q(t)$ is the outflow [$L^{-1}T$] at time t, a is the storage coefficient [$L^{1-nl}T^{nl}$], and nl is the coefficient of non-linearity [-].

Additionally, a mass balance equation describes the rate of change in storage ($ds(t)/dt$) as the difference between inflow ($u(t)$ [$L^{-1}T$]) and outflow ($q(t)$ [$L^{-1}T$]),

$$\frac{s(t)}{dt} = u(t) - q(t) \tag{3.36}$$

Setting $nl = 1$ leads to the most widely used form of the conceptual reservoir, the linear reservoir. In this case the parameter a represents the residence time T [T], and the outflow of the reservoir is directly proportional to the storage content. Combining the storage function and the mass balance equations yields the following model formulation,

$$\frac{dq}{dt} = \frac{1}{T}[u(t) - q(t)] \tag{3.37}$$

The linear reservoir is functionally identical to the following first-order discrete-time transfer function (TF) (e.g., Lees, 2000),

$$q_t = \frac{b_0}{1 - a_1 z^{-1}} u_t \; ; \; b_0 = \frac{\Delta t}{T} \; ; \; a_1 = 1 - \frac{\Delta t}{T} \tag{3.38}$$

where Δt is the discretization step size, and z^{-i} is the backward shift operator defined as

$$z^{-i} q_t = q_{t-i} \tag{3.39}$$

Chow *et al.* (1988) show how a general form of the TF model can be derived from a general hydrologic system model (Chow and Kulandaiswamy, 1971). The general form of an n^{th} order single input single output (SISO) discrete time-system can be written as

$$q_t = \frac{B(z^{-1})}{A(z^{-1})} u_{t-\delta} \tag{3.40}$$

where δ represents a lag element, and $A(z^{-1})$ and $B(z^{-1})$ are the following polynomials,

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_n z^{-n} \tag{3.41}$$

and

$$B(z^{-1}) = b_0 + b_1 z^{-1} + \cdots + b_m z^{-m} \tag{3.42}$$

The TF model structure is described by the triad $[n,m+1,\delta]$. This model can represent any combination of linear reservoirs connected in parallel and/or series as shown in Fig. 3.6, using partial fraction expansion to perform the decomposition (Young, 1992).

Advantages of the TF representation of a system of linear conceptual reservoirs include the ability to apply powerful system identification techniques for optimal parameter estimation, and increased structural flexibility (Lees, 2000). A system identification technique that has been successfully used to identify and estimate TF models in the context of RR modelling is the Simple Refined Instrumental Variable technique (SRIV, Young, 1992; Young *et al.*, 1996). The TF identification software

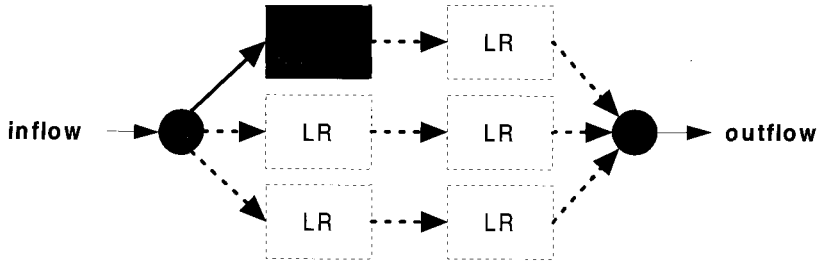


Fig. 3.6 General linear reservoir model.

used in the RRMT is based closely on that available in the CAPTAIN toolbox (http://www.es.lancs.ac.uk/cres/captain/), which was developed by Peter Young and colleagues at the University of Lancaster. Details of the approach can be found in Young (1984), and Young and Jakeman (1979a, 1979b, 1981).

The conceptual reservoir can also be used in RRMT in its non-linear form ($nl \neq 1$). Linear and non-linear reservoirs can additionally be combined in parallel or serial structures.

The routing components described so far are based on the assumption that all subsurface runoff drains from the catchment via the stream. However, part of the runoff might leave some catchments through subsurface pathways. A routing component taking this into account (here called leaky catchment structure) can consist of a linear reservoir with different outlets (Chapman, 1999; Sugawara, 1995; Moore, 1999). The outflow from the bottom outlet represents the part of the effective rainfall not contributing to streamflow (Fig. 3.7c). The leaky catchment structure can be described in numerical form as follows.

The catchment losses, from the bottom outlet, are

$$q_3 = k_3 s \tag{3.43}$$

where k_3 [T^{-1}] is a time constant and s[L] is the storage content. The contribution to catchment runoff q[LT^{-1}] is described by

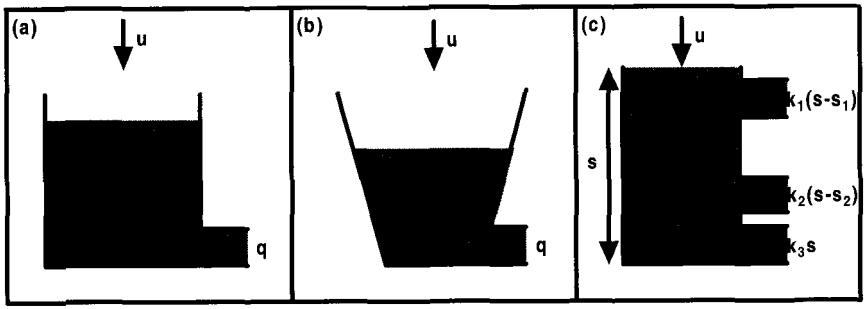


Fig. 3.7 Conceptual routing components available in the RRMT: (a) linear; (b) non-linear (example); and (c) leaky catchment structure.

$$q = \begin{cases} 0 & ,if\ s \le s_2 \\ k_2(s - s_2) & ,if\ s_2 < s \le s_1 \\ k_2(s - s_2) + k_1(s - s_1) & ,if\ s > s_1 \end{cases} \tag{3.44}$$

with k_1 [T^{-1}] and k_2 [T^{-1}] being time constants for the different outlets, and s_1 [L] and s_2 [L] as the corresponding thresholds.

3.2.5 Optimization modules

The model structures that can be implemented in the RRMT contain parameters that typically refer to a collection of aggregated processes. Therefore they often do not have a direct physical interpretation and cannot be measured in the field. Instead they can be estimated using a calibration procedure whereby the model parameters are adjusted until system and model output show an acceptable level of agreement. The agreement is typically measured using an objective function (OF), *i.e.*, some measure of the aggregated size of the model residuals that represent the differences between observed and calculated system output, supported by visual inspection of the calculated time series. The model that produces the best agreement is commonly assumed to be representative of the natural system under investigation.

Most parameters in conceptual rainfall-runoff models define non-linear model equations; *i.e.,* they enter the model equations in a non-linear way. The consequence of this is that an iterative search is required to identify the optimum parameter values (a more detailed discussion appears in Chapter 2). This can be done using a manual *trial-and-error* procedure, an automatic search algorithm, or a combination of both approaches (Boyle *et al.*, 2000; Hogue *et al.*, 2000). Manual calibration is time-consuming and difficult to achieve in the presence of parameter dependence. Automatic calibration algorithms have the potential to reduce this problem. Available search algorithms can be separated into local and global approaches. Local search algorithms start from an initial solution, *i.e.,* an initial parameter set, and try to sequentially improve this solution by repeatedly moving through the parameter space using various schemes to find the next location. The search is stopped when a termination criterion, e.g., a specific OF value, is satisfied. Research has shown that the characteristics of the response surface created by conceptual rainfall-runoff models are usually not suitable for application of local optimization methods (Duan *et al.*, 1992), since the presence of multiple optima often leads to premature convergence of the optimization process at a local optimum. Global optimization methods, working with parameter or model populations (*i.e.,* parameter set/model structure combinations), have therefore been introduced. Popular approaches include population evolution (e.g., Wang, 1991; Duan *et al.*, 1992) and adaptive random search methods (e.g., Price, 1987).

The SCE (Duan *et al.*, 1992, 1993, 1994; Sorooshian *et al.*, 1993; Hogue *et al.*, 2000) is reputedly the most successful approach with respect to the calibration of hydrological model structures (e.g., Gan and Biftu, 1996; Kuczera, 1997; Thyer *et al.*, 1999). The general algorithm contains the following steps:

- A population of models (parameter sets within a selected model structure) is randomly sampled from the n-dimensional feasible parameter space.
- The population is divided into complexes, each containing $2n + 1$ models.
- Each complex is evolved independently using a deterministic downhill simplex. Models to form the simplex are selected based on a

probability distribution which favours more promising models (*i.e.*, those that produce a better OF value).

- The models of all complexes are mixed again after each complex has been evolved individually, *i.e.*, a shuffling step has been performed. This ensures that information is shared and is probably the main feature that makes the SCE so successful.
- New complexes are formed and the evolution step is repeated. This iteration continues until a specified termination criterion is satisfied.

Various researchers have found that different parameter sets, often widely distributed in the feasible parameter space, can lead to similar model performance with respect to a certain OF (e.g., Johnston and Pilgrim, 1976; Beven and Binley, 1992). As discussed in greater detail in Chapter 2, these findings have led some researchers to suggest that the idea of *point identifiability*, *i.e.*, a global optimum, is not appropriate in the presence of errors in data and model structure or limitations in parameter estimation procedures (e.g., Spear and Hornberger, 1980; Van Straten and Keesman, 1991; Beven and Binley, 1992). Instead, these researchers advocate the identification of *behavioural* parameter populations, e.g., parameter sets that perform better than a certain threshold performance value. All parameter sets that are classified as behavioural are considered to be possible representations of the natural system under investigation (Van Straten and Keesman, 1991). Monte-Carlo sampling procedures (Press *et al.*, 1992) such as importance sampling or Markov chain sampling (e.g., Kuczera and Parent, 1998) are sometimes used as a basis for exploring the feasible parameter space in order to identify potential models. However, pure random sampling based on a uniform prior distribution is most commonly applied (e.g., Beven and Binley, 1992; Freer *et al.*, 1996). This method assumes no prior knowledge of parameter values other than minimum and maximum values. A drawback of this sampling approach is its requirement for a large number of model runs.

Gupta *et al.* (1998) pointed out that in the presence of unavoidable model structural error, a range of parameter sets – optimal with respect to different OFs – is required to adequately simulate all response modes of a natural system. Single parameter sets will favour specific response features, e.g., peaks or low flows. This leads to the conclusion that a multi-

objective optimization problem exists even for single output mode structures. The implications of this observation are discussed further below.

In those cases where a TF model is selected as the routing component, structure identification and parameter estimation is performed using the SRIV method of system identification (Young, 1992; Young *et al.*, 1996). Since the TF model is linear, the standard least squares normal equations can be used to calculate the optimum parameter estimates under a number of assumptions relating to the form of the random inputs to the system. The TF structure is identified by fitting a large number of different model structures and the subsequent assessment of the model performance versus parameter identifiability using an extension of the Akaike Information Criterion (AIC; Akaike, 1974) termed Young's Information Criterion (YIC; Young *et al.*, 1996). This statistical assessment is combined with an assessment of the physical validity of the model in an approach termed *data-based mechanistic modelling* (Young, 1992; Young and Beven, 1994; Young *et al.*, 1996; Lees, 2000). In the RR modelling case this means that a TF structure is only accepted if the TF structure can be interpreted in a physical way (Young, 1992), *i.e.*, as a combination of linear stores in series and/or parallel with parameters describing aspects such as residence times or fractions of flow contributing to particular pathways.

In summary, the RRMT contains a version of the SCE and SRIV. A Monte Carlo sampling procedure is also available. Random sampling based on uniform prior distribution is usually appropriate for the parsimonious models implemented in the RRMT.

3.2.6 Objective functions

The performance of a model is typically judged using an OF, usually in combination with visual inspection of the calculated hydrograph. OFs aggregate the model residuals, *i.e.*, the part of the observed flow not reproduced by the model, which can be calculated as

$$\varepsilon_k(\theta) = y_k - \hat{y}_k(\theta) \tag{3.45}$$

where $\hat{y}_k(\theta)$ is the calculated flow at time-step k using the parameter set θ, y_k is the observed flow at time-step k, and $\varepsilon_k(\theta)$ is the resulting residual at time-step k using parameter set θ.

The task is then to minimize or maximize the size of the OF (depending on definition). A variety of functions are available in the RRMT (Table 3.1), which can be used to evaluate different aspects of a model's performance.

The most commonly utilized OFs in hydrological modelling are variations of the Simple Least Squares (SLS) function,

$$SLS(\theta) = \sum_{k=1}^{N} \varepsilon_k(\theta)^2 \tag{3.46}$$

where N is the number of flow values available. The SLS function is the maximum likelihood estimator when the following assumptions about the residuals cannot be rejected (Troutman, 1985; Yapo *et al.*, 1996; Gershenfeld, 1999): (1) the residuals are independent and identically distributed (i.i.d.), (2) the residual distribution has homogeneous variance, and (3) the residuals are normally distributed with zero mean.

The analysis of the characteristics of the residual distribution can also be an important step in evaluating the suitability of a model structure (e.g., Beck, 1981a, 1981b; Yapo *et al.*, 1996; Mroczkowski *et al.*, 1997). "If a fit produces residuals consistent with the random error assumptions, then the model has extracted all useful information from the data leaving only noise in the residuals" (Mroczkowski *et al.*, 1997).

Graphical tests can be applied to evaluate the assumptions made about the characteristics of the residuals (Draper and Smith, 1981; Kuczera, 1983). A number of different plots are available in the RRMT to facilitate this evaluation (Fig. 3.8): (1) plotting the residuals versus predicted and calculated runoff reveals whether the variance of the residuals increase with increasing flow values, *i.e.*, the problem of heteroscedascity; (2) plots of residuals versus time reveal long term effects (trends) or dependency in time; (3) frequency distribution plots, also showing the mean, can be used to indicate how close the residual distribution is to a normal distribution; and (4) calculating (and plotting) the autocorrelation coefficients allows users to assess the correlation of the residuals in time.

If the assumption of zero mean cannot be rejected, the autocorrelation coefficient of the residuals ac is described (Scholz, 1995) as:

$$ac(\tau) = \frac{\sum_{k=1}^{N-\tau} \varepsilon_k(\theta) \cdot \varepsilon_{k+\tau}(\theta)}{\sum_{k=1}^{N-\tau} \varepsilon_k^2(\theta)} \tag{3.47}$$

where τ is a lag ($\tau = 0,1,2,\ldots,N\text{-}1$), and $ac(\tau)$ ranges from minus one to plus one.

The 95% confidence intervals can be calculated as follows,

$$| ac(\tau) | < 1.96 \cdot \frac{1}{\sqrt{N}} \tag{3.48}$$

Not more than 5% of the residuals should lie outside these limits.

The assumption of no autocorrelation, *i.e.,* independence in time, is often not satisfied in RR modelling applications. Residuals of small (e.g., daily) temporal discretization are usually related over a number of time-steps. Sorooshian and Dracup (1981), and Kuczera (1983) describe consequences and possible corrective measures if this assumption is violated.

Another problem often encountered in RR modelling is the fact that the residual variance increases with increasing flow values; *i.e.,* the assumption of homoscedascity cannot be justified (Sorooshian and Dracup, 1981). In such cases, the variance can be stabilized through transformation of the simulated and observed flow data, or by the use of a weighted least-squares OF (Kottegoda and Rosso, 1997).

A Box-Cox transformation (Box and Cox, 1964), which can be written in the following form, is a useful transformation in this regard,

$$q^*(\lambda) = \begin{cases} \dfrac{q^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log(q) & \text{for } \lambda = 0 \end{cases} \tag{3.49}$$

where q is the original flow, q^* is the transformed flow and λ is the transformation parameter.

Table 3.1 Some of the objective functions available in the RRMT.

Objective function	Notation	Equation
Nash-Sutcliffe Efficiency[1]	NSE	$NSE(\theta) = 1 - \frac{\sum_{k=1}^{N} (y_k - \hat{y}_k(\theta))^2}{\sum_{k=1}^{N} (y_k - \bar{y})^2}$
Root Mean Square Error	RMSE	$RMSE(\theta) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y_k - \hat{y}_k(\theta))^2}$
Heteroscedastic Maximum Likelihood Estimator[2]	HMLE	$\min_{\theta,\lambda} HMLE = \frac{\frac{1}{N} \sum_{k=1}^{N} w_k (y_k - \hat{y}_k(\theta))^2}{\left[\prod_{k=1}^{N} w_k \right]^{1/N}}$ $w_k = y_k^{2(\lambda-1)}$
Bias	BIAS	$BIAS(\theta) = \sum_{t=1}^{N} (y_k - \hat{y}_k(\theta))$
Deviation of Runoff Volumes[3]	DV	$DV(\theta) = \frac{\sum_{k=1}^{N} \hat{y}_k(\theta)}{\sum_{k=1}^{N} y_k}$
RMSE Response Modes[4]	FD,FQ,FS	RMSE of selected time steps
RMSE Horizontal Segmentation	FH,FM,FL	RMSE of selected time steps
RMSE Warming Up Period	FWU	RMSE of selected time steps

[1] Nash and Sutcliffe (1970)
[2] Sorooshian and Dracup (1980), Yapo *et al.* (1996)
[3] ASCE (1993)
[4] Boyle *et al.* (2000)

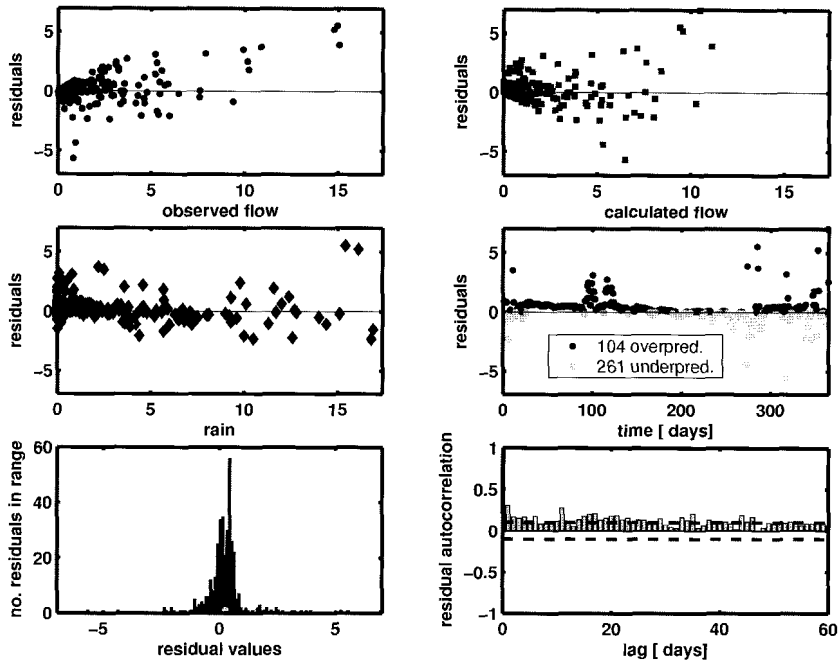


Figure 3.8 Residual analysis plots.

Sorooshian and Dracup (1981) use this transformation in their Heteroscedastic Maximum Likelihood Estimator (HMLE, Table 3.1, see also Sorooshian *et al.*, 1993). The parameter λ is optimized every time the OF is calculated. The HMLE is the maximum likelihood estimator when assumptions (1) and (3) are valid, and the residual distribution has mean zero and heteroscedastic variance, *i.e.*, the variance is a function of the flow. The Box-Cox transformation and the use of the HMLE are both available options in the RRMT.

Gupta (2001) points out that the analysis of the above outlined aspects is based on statistical regression analysis. How far these assumptions need to be satisfied for a dynamic conceptual model to perform well may therefore be questionable. For example, the residuals will commonly be correlated in time in this type of model, but does this violation matter with respect to the physical realism of the parameter esti-

mates for any practical purposes? Further research is required to answer the question of the importance of this type of model testing.

Another problem, already briefly indicated, occurs in terms of selection of an appropriate OF when automatic search algorithms are applied. The drawback of single-criterion algorithms is that the calibration result is fully dependent on one OF (Gupta *et al.*, 1998; Boyle *et al.*, 2000). This can lead to an overemphasis on a certain aspect of the response, e.g., peak flows, while neglecting the model performance with regard to another aspect, e.g., low flows (Fig. 3.9). Hydrological models are typically not capable of fitting all system response modes with a single parameter set due to the presence of structural errors. This can lead to a calibration result that is not acceptable to hydrologists, thereby limiting usefulness of automatic calibration (Boyle *et al.*, 2000). A multi-criteria approach to address this problem is proposed by Gupta *et al.* (1998). The objective is to increase the amount of information retrieved from the model residuals to (1) find the parameter population necessary to fit all aspects of the observed output time series, e.g., in a first stage of a hybrid automatic-manual calibration procedure (Boyle *et al.*, 2000); (2) increase the identifiability of the model parameters (Wagener *et al.*, 2001); and (3) assess the suitability of the model structure to represent the natural system, *i.e.*, to identify model structural insufficiencies (Gupta *et al.*, 1998; Boyle *et al.*, 2000).

A simple application of multi-objective optimization is the definition of specific OFs for water resource management. A specific flow range (e.g., between the minimum environmentally acceptable flow and a maximum water supply abstraction rate) is often of particular interest. Specifying an OF measuring the performance of the model in this range in addition to the traditional measures of performance, can help to assess the suitability of a model structure for the selected purpose (Lees and Wagener, 2000a, 2000b). Measures available in the RRMT allow the calculation of the RMSE above and below user-specified thresholds, and within a certain flow range (see Table 3.1).

Other multiple objectives that can be selected within RRMT are the RMSE based on a segmentation scheme modified from Boyle *et al.* (2000; see application example in Section 3.4 for details), or the RMSE for a warming up period to estimate initial conditions, FWU. The RRMT

allows users to add OFs in a modular way depending on an application's requirements.

3.2.7 Visual analysis modules

Various plotting options are available in the RRMT to analyse the data and the performance of the model. Examples of these plots are: (1) double mass plots; (2) observed versus calculated flow scatter diagrams, both on normal and logarithmic scales; and (3) flow duration and volumetric fit curves, etc. These plots are helpful tools that enable an assessment of model performance from different perspectives.

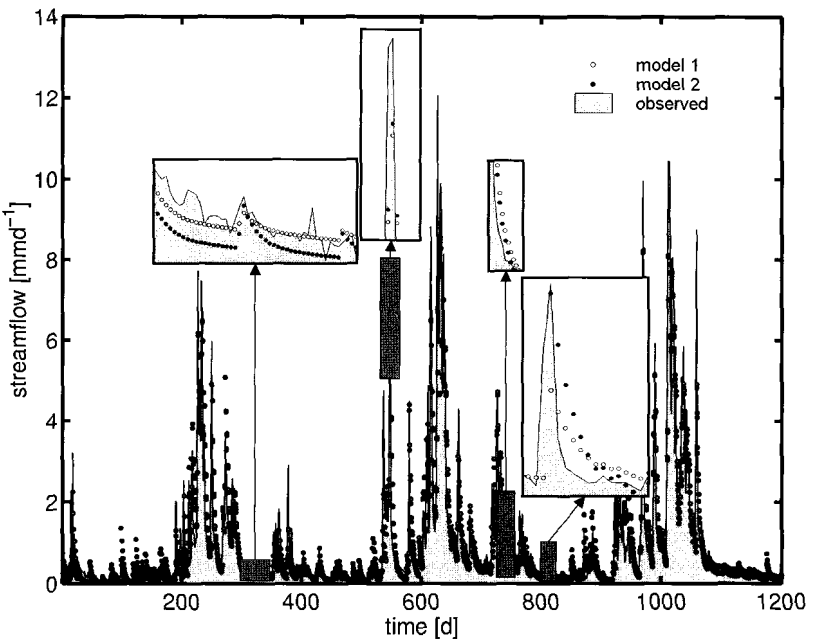


Fig. 3.9 Plot showing predictions using two different parameter sets with an identical model structure. Both realizations yield similar values of the NSE measure (0.82), but show differences in fit when the response is analysed closely. (Reproduced from Wagener *et al.*, 2002a; with permission of Water Resources Publications LLC.)

3.2.8 Summary

In summary, the RRMT is a generic modelling shell that allows the user to implement, evaluate, and modify lumped, parametric, or hybrid metric-parametric model structures. A variety of moisture accounting and routing components are provided and the addition of new structural elements is straightforward. Different data manipulation, optimization, and visualization options are available to calibrate and evaluate the model.

3.3 Monte Carlo Analysis Toolbox

3.3.1 General

Increasingly important parts of any modelling exercise are the detailed investigation of model performance in terms of parameter sensitivity and identifiability, the suitability of a particular model structure, and prediction uncertainty. Understanding model behaviour and performance increases the transparency of the modelling procedure and helps in the assessment of the reliability of modelling results.

The Monte Carlo Analysis Toolbox (MCAT, Wagener *et al.*, 1999; 2002a) includes a number of analysis methods to evaluate the results of Monte Carlo parameter sampling experiments or model optimization methods based on population evolution or random techniques. Functions contained in the MCAT include an extension of the Regional Sensitivity Analysis (RSA, Spear and Hornberger, 1980; Hornberger and Spear, 1981) proposed by Freer *et al.* (1996), various components of the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992; Freer *et al.*, 1996), and options for the use of multiple-objectives for model assessment (Gupta *et al.*, 1998; Boyle *et al.*, 2000), response surface plots. It also contains a novel Dynamic Identifiability Analysis (DYNIA) approach.

3.3.2 System architecture

The MCAT is a collection of MATLAB (Mathworks, 1996) analysis and visualization functions integrated through a GUI (Fig. 3.10). It can be accessed through an interface from the RRMT. Note, however, that the MCAT is not specifically related to RR modelling and can be used to analyse the results of any dynamic mathematical model (off-line).

3.3.3 Parameter sensitivity and identifiability

Sensitivity analysis is an approach to evaluating how changes in model parameters affect the model output variable(s). This information can be used to identify parameters that are not important for the reproduction of the system response and can therefore be subsequently fixed or removed, reducing the dimensionality of the calibration problem.

A popular sensitivity analysis method that utilizes the results of Monte Carlo sampling is Regional Sensitivity Analysis (RSA, Spear and Hornberger, 1980; Hornberger and Spear, 1981), which analyses the sensitivity of the model output to changes in the parameters without refer-

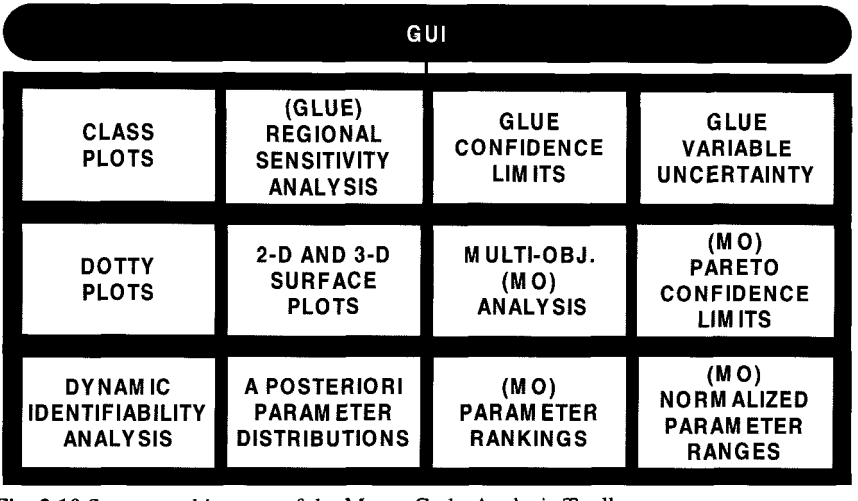


Fig. 3.10 System architecture of the Monte Carlo Analysis Toolbox.

ring to a certain point in the parameter space, such as the most likely value for a specific parameter (Spear, 1993).

The RSA method starts with a Monte Carlo sampling of N points in the feasible parameter space, drawn from a multivariate uniform distribution. The sampled parameter population is partitioned into a behavioural (B) and a non-behavioural ($\overline{B}$) group. By behavioural we mean parameter sets that produce a model response (behaviour) similar to the response of the hydrological system. The division into behavioural and non-behavioural can, for example, be based on the predicted state of the system (e.g., Spear and Hornberger, 1980) or on a measure of performance (e.g., Hornberger *et al.*, 1985; Beven and Binley, 1992). The cumulative distributions for the two groups ($F(\theta_x \mid B)$ and $F(\theta_x \mid \overline{B})$) are computed (Fig. 3.11). A separation between the distribution curves indicates a statistical difference between the characteristics of the two (behavioural and non-behavioural) subpopulations. This indicates that the parameter θ_x is sensitive, *i.e.*, its value can be strongly correlated with model performance. The significance of the separation can be estimated using statistical tests such as the Kolmogorov-Smirnov (KS) two-sample test (Kottegoda and Rosso, 1997), and a heuristic ranking scheme can be in-

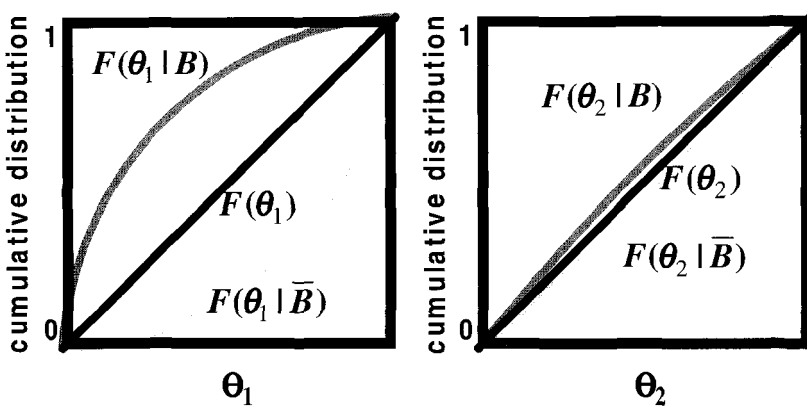


Fig. 3.11 Cumulative distributions of initial ($F(\theta_x)$), 'behavioural' ($F(\theta_x \mid B)$) and 'non-behavioural' ($F(\theta_x \mid \overline{B})$) populations for a sensitive parameter θ_1, and a (conditionally) insensitive parameter θ_2.

troduced based on the actual values of the KS measure (Spear and Hornberger, 1980).

An unfortunate weakness of this approach is that a lack of separation between the cumulative distributions of $F(\theta_x \mid B)$ and $F(\theta_x \mid \overline{B})$ is only a necessary, and not a sufficient condition for insensitivity of θ_x (Spear, 1993). It can also be caused by strong correlation with other parameters. Evaluation of the parameter covariance can be used to estimate whether this is the case (Hornberger and Spear, 1981; Hornberger *et al.*, 1985). The interaction between two parameters can also be investigated in the MCAT by plotting their response surface with respect to a particular OF.

A modification of the RSA approach proposed by Freer *et al.* (1996) is implemented in the MCAT to support visual inspection of the sensitivity of the different parameters with respect to a selected OF (Fig. 3.12). Freer *et al.* (1996) split the parameter population, ranked on the basis of their OF values, into ten groups of equal size and plot the cumulative distribution of the parameters in each group with respect to the chosen measure of performance. Differences in form and separation of the resulting curves provide information about parameter sensitivity. Splitting

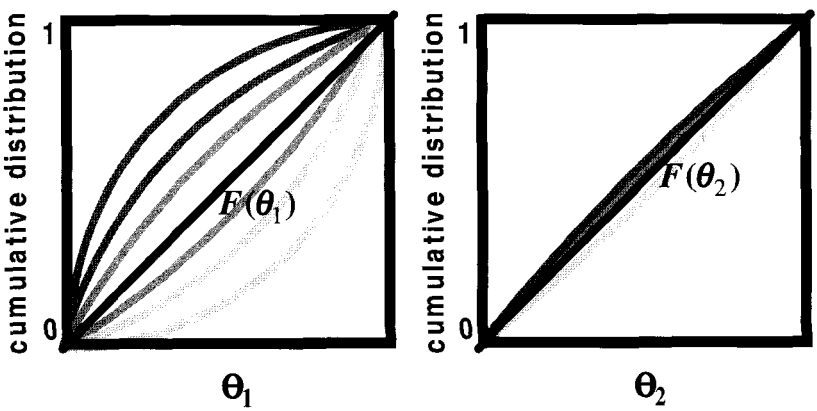


Fig. 3.12 Schematic Regional Sensitivity Analysis plot using the modified version of Freer *et al.* (1996). A sensitive parameter is shown in the left plot, while the right one shows an insensitive one.

the parameter population into ten groups, instead of merely dividing it into behavioural and non-behavioural parameter sets as in the original method, avoids the need to (subjectively) select a threshold value, and increases the information provided by the analysis.

Variation in performance between the different groups can be visualized using the class plot option in the MCAT. This figure shows the response vector calculated with the best performing parameter set in each of the groups and plots them together with the observed response (if available).

Parameter sensitivity is a necessary, but not a sufficient requirement for identifiability, since values of sensitive parameters that produce an acceptable model performance can still be distributed over a relatively wide range of the feasible parameter space. A model, *i.e.*, a parameter set θ within a certain model structure, is termed (globally) identifiable if it is possible to uniquely determine its location in the parameter space, based on the model output produced. This requires the parameter set to yield a unique response vector (Mous, 1993). However, the specific characteristics of parametric RR models (Duan *et al.*, 1994; Gupta *et al.*, 1998), the often limited information content of the available time-series, and the restrictions of single value OFs that aggregate the fit to the response vector into a single value limit the success of parameter identification in many cases, as described in detail in Chapter 2.

The RSA procedure outlined above has been extended to investigate the identifiability of a parameter (Wagener *et al.*, 2001). Reducing the analysis to the cumulative distribution of the best performing group derived from the approach implemented by Freer *et al.* (1996) allows the definition of an empirical measure of identifiability for each parameter. The cumulative distribution of a uniform distribution is a straight line. Deviations from this straight line indicate regions of higher identifiability. Splitting the feasible parameter range into segments and calculating the gradient of the cumulative distribution in each segment leads to an indicator of identifiability. Figure 3.13 shows how gradients for a well identified and a poorly identified parameter are distributed. An example of the utility of this measure is included as part of the demonstration application described in Section 3.4. The identifiability measure can also be used in a dynamic way to link parameters (and related model compo-

nents) to system response modes in an objective fashion. This technique will be introduced in Chapter 4.

3.3.4 Model structure suitability

The problem of model structural error or suitability is an issue of increasing interest in recent research (e.g., Kleissen *et al.*, 1990; Yapo *et al.*, 1996; Gupta *et al*, 1998; Boyle *et al.*, 2000). Gupta *et al.* (1998) introduced a multi-objective analysis framework to investigate deficiencies in the model structure, which are reflected in a structure's inability to simultaneously reproduce different aspects of the system response with a single set of parameters. Different parameter combinations are required to fit different response modes (Boyle *et al.*, 2000) or output variables (Bastidas, 1998). Defining more than one OF to measure the model performance with respect to different system responses, and analysing the variation in the resulting parameter populations can be used as an indicator of the degree of structural deficiency present. Multi-objective plots, *i.e.,* scatter plots of different OFs against each other, are available in the MCAT to evaluate whether the selected measures of performance retrieve similar information from the residuals or whether they are uncorrelated. Uncorrelated measures often show a trade-off front, which can be described by a Pareto set, as shown by Gupta *et al.* (1998). A parameter set is termed *Pareto optimal* if, by changing its parameters, an improvement with respect to one OF results in the degradation with respect to another OF (Chankong and Haimes, 1993). An estimate of the Pareto set can be calculated from the available parameter population in the MCAT and the range of predictions produced by this population can be visualized. It is also possible to plot the values of the best parameter sets as well as their predictions (albeit not simultaneously) with respect to different OFs on a normalized scale to see the variation associated with the different measures.

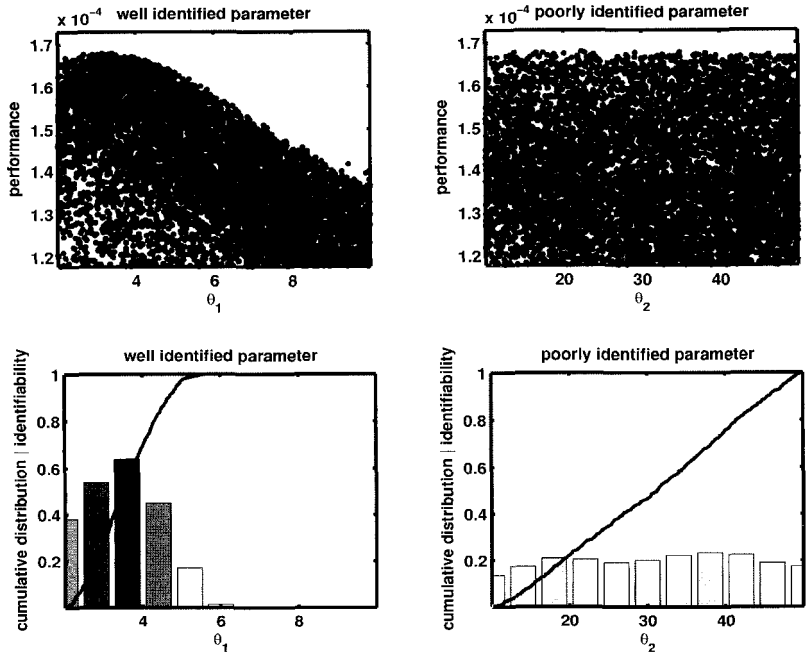


Fig. 3.13 Example of a well-identified parameter (left column) and a poorly identified parameter (right column). The top row shows scatter plots of parameter versus measure of performance. It has to be considered that these projections into a single parameter dimension can, however, hide some of the structure of the response surface (Beven, 1998). The bottom row shows the cumulative distribution of the best performing 10% of parameter sets and the corresponding gradients within each segment of the parameter range.

3.3.5 Prediction uncertainty

Uncertainties in the model parameters are propagated into the model predictions. A popular method of estimating and propagating this uncertainty is the Generalized Likelihood Uncertainty Estimation (GLUE) approach (Beven and Binley, 1992; Freer *et al.*, 1996; Beven, 1998). The underlying assumption of this approach is similar to that of the RSA methodology described earlier. Because it is not possible to find a global optimum, only a population of possible (behavioural) parameter sets can be estimated. The approach starts with a Monte Carlo procedure, sampling a large number of random parameter sets (usually) from uniform

distributions, assuming only a minimum amount of prior information is available (upper and lower bounds). A simulation is then performed using each parameter set and a *likelihood* measure to evaluate the performance of each set is calculated. A likelihood measure in the GLUE context can be any measure of performance so long as better performing models attain a higher value and the sum taken over all sampled parameter sets is unity. The likelihood values of parameter sets below a certain user selected threshold are set to zero, *i.e.,* these are considered to be non-behavioural. Likelihood values of the retained models are used to weight the corresponding predictions (Beven, 2000a, p.236). Uncertainty in the parameter values, and (implicitly) the uncertainty in the data are propagated into the model response in the form of confidence limits at a specified percentile. Within the MCAT, it is also possible to calculate a cumulative probability distribution and probability density function for a selected variable, e.g., the peak output.

The MCAT also contains a plot that shows the range of predictions produced by the Pareto set from a multi-objective analysis. This is the trade-off range caused by the inability of a single parameter set to reproduce all response modes. This prediction range is usually narrower than the one produced by GLUE (see discussion in Chapter 2).

3.3.6 Summary

In summary, the MCAT aims to provide a tool that addresses the emerging requirement to include detailed investigations of model behaviour, performance, and prediction uncertainties as an integral part of the modelling process. Options currently available include visualization of the parameter space (e.g., surface and identifiability plots), various capabilities of the GLUE approach (e.g., prediction uncertainty and extended RSA approach), and plots to analyse the role of multiple-objectives (including the option to calculate the Pareto optimal solution and plot the response range produced by it). It also allows the estimation of dynamic parameter identifiability, an approach introduced in Chapter 4.

3.4 Application Example — Complexity Versus Identifiability

The following application example is presented to illustrate the use of both toolboxes using data from the River Medway catchment (1256.1 km^2), located in southeast England. Almost seven years of daily naturalized flows, rainfall, PE and temperature data are available for use in this modelling exercise.

The Medway catchment is characterized by a mixture of permeable (chalk) and impermeable (clay) geologies subject to a temperate climate (annual rainfall of 772 mm and an annual PE of 663 mm over the period 1989-96).

Different response modes are identified using a partitioning scheme suggested by Boyle *et al.* (2000) but adapted for this particular catchment. This approach is based on the assumption that the behaviour of the catchment is different during periods *driven* by rainfall and periods without rain, *(non-driven)*. Further, the periods classified as non-driven can be additionally discretized between periods dominated by quick response and periods dominated by the slow response of the catchment system. The streamflow hydrograph can therefore be partitioned into three components: *driven*, *non-driven quick*, and *non-driven slow*.

The time-steps that correspond to each period are identified through an analysis of the precipitation data and the time of concentration of the catchment (Fig. 3.14). Time-steps with rainfall larger than a certain threshold (e.g., the mean of the square-root of the rainfall), lagged by the time of concentration of the catchment, are classified as driven, Q_D. Of the remaining (non-driven) time-steps, those with streamflow lower than a selected threshold (e.g., a third of the mean of the square-root of the flows) are classified as non-driven-slow, Q_S, and the rest are classified as non-driven-quick, Q_Q. Model performance during these three periods is estimated by calculating an OF using the residuals during each period.

The performance of each model structure is evaluated using the Nash-Sutcliffe Efficiency measure (NSE, Nash and Sutcliffe, 1970, Table 3.1). However, the NSE criterion favours models that are able to reproduce the catchment response at high flows (Legates and McCabe, 1999). Improvement in performance through increased model complexity, for ex-

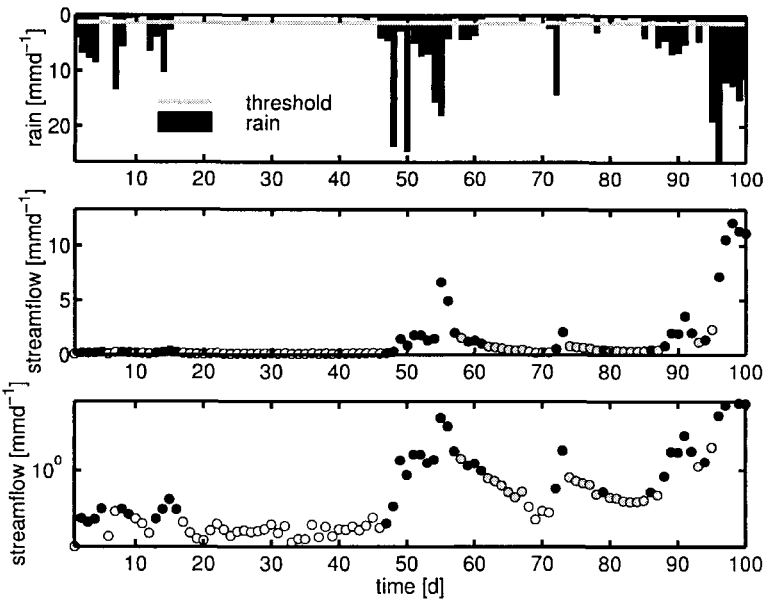


Fig. 3.14 A hundred day period of Medway streamflow and rainfall segmented using a scheme modified from Boyle *et al.* (2000). The dots indicate time steps classified as driven (black), non-driven quick (grey), and non-driven slow (white). (Reproduced from Wagener *et al.*, 2002a; with permission of Water Resources Publications LLC.)

ample during periods of recession, could therefore go unnoticed. Using the partitioning scheme introduced above, and calculating the weighted addition of the NSE for each period can avoid this. Equal weight is given to each of the periods, assuming that a balanced reproduction of the whole hydrograph is the objective. The measure of performance is calculated as follows:

$$PEF = \frac{1}{3} NSE_D + \frac{1}{3} NSE_Q + \frac{1}{3} NSE_S \tag{3.50}$$

where NSE_x is the performance measure during driven (Q_D), non-driven quick (Q_Q), and non-driven slow (Q_S) periods. A MATLAB implementation of the SCE approach (Duan *et al.*, 1992) is used to optimize the model performance with respect to the defined OF, *PEF*.

Table 3.2 Model structures used. A cross indicates the use of this particular structural element.

Model	Moisture accounting			Routing Type		Structure		
	B[1]	P2S[2]	BP[3]	L[4]	NL[5]	1S[6]	2P[7]	3P[8]
01 (2)[0]	X			X		X		
02 (2)		X		X		X		
03 (3)		X	X	X		X		
04 (4)		X	X		X	X		
05 (5)		X	X	X			X	
06 (6)		X	X	X	X		X	
07 (7)		X	X	X				X

[0] Number of parameters.
[1] Simple bucket emptying at potential rate as long as moisture is available.
[2] Penman two store model as described earlier in the text (bypass = 0).
[3] Bypass mechanism, *i.e.*, a percentage of the rainfall is directly contributing to runoff.
[4] Conceptual linear reservoirs.
[5] Conceptual non-linear reservoirs.
[6] Single reservoir.
[7] Two reservoirs in parallel.
[8] Three reservoirs in parallel.

The identifiability of the model parameters is investigated using the empirical measure introduced in Section 3.3.3. A Monte Carlo sampling procedure (10,000 samples) based on uniform distributions is used to investigate the feasible parameter space for each model structure. The maximum identifiability value of each parameter, *i.e.*, the maximum gradient of its cumulative distribution split into bins, is calculated from the best performing 10% parameter values. The mean of these maximum identifiability values for each model structure is used as an indicator of parameter identifiability. The difference in parameter dimension between the model structures is considered by analysing the stability of the identifiability measure while varying the size of the parameter population used to calculate it.

The graph shown in Fig. 3.15 depicts measures of performance and identifiability plotted against model structures. The result suggests that increased model complexity leads to increased performance while the identifiability of the model parameters decreases. A trade-off between

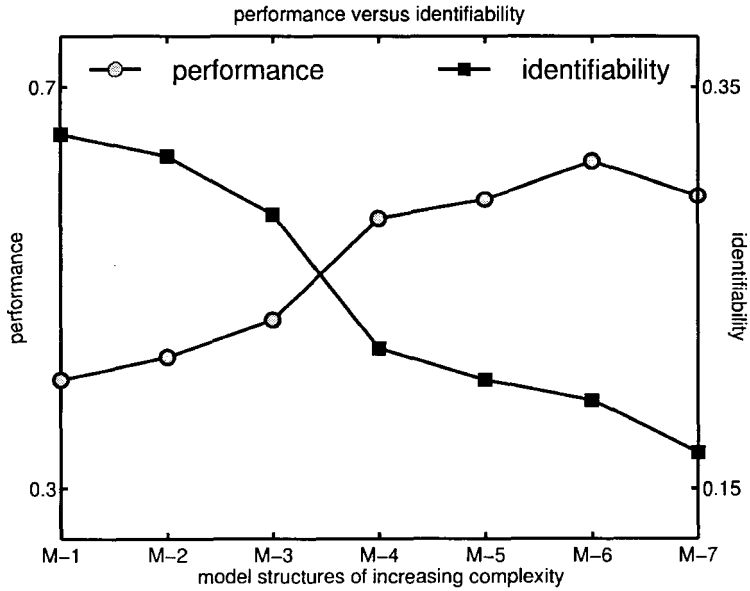


Fig. 3.15 Performance versus identifiability plot for the different model structures investigated. (Reproduced from Wagener *et al.*, 2002a; with permission of Water Resources Publications LLC.)

these two features is required if, for example, it is intended that the model structure be used within a regionalization project. Note that the decline in performance from model M-6 to model M-7 is caused by the less suitable slow flow component of model M-7. Further research has shown that other segmentation schemes are better suited for catchments in similar climatic regions (see Chapter 4).

3.5 Summary and Conclusions

A toolkit has been implemented that facilitates the development and analysis of lumped and parsimonious model structures, using state-of-the-art modelling techniques.

The RRMT allows the implementation of parametric or hybrid metric-parametric model structures. Its major strength is a high degree of structural flexibility which allows the quick implementation and evalua-

tion of different model structures to identify the most suitable one(s) for the circumstances at hand. The option to run the RRMT in batch mode is particularly attractive for studies that include a range of model structures or a large number of catchments.

The MCAT enables the detailed investigation of model performance, parameter sensitivity and identifiability, model structure suitability, and prediction uncertainty. It is generic and can be applied to the analysis of any dynamic mathematical model structure. Within the MCAT, a new empirical identifiability measure has been implemented, based on the gradient of the cumulative parameter distribution.

A brief application example was used to show how the toolkit can be applied. Analysis of a variety of model structures with different levels of complexity showed that the model performance increased with increasing complexity, while parameter identifiability decreased. This demonstrates that there is a clear trade-off between these two features. The study used a MO approach, assessing different components of hydrograph response, to ensure that any improvement in performance did not go unnoticed.

The latest version of the toolboxes and full manuals are available from the website of the Environmental and Water Resource Engineering Section of the Imperial College London. For details see Appendices B and C.

This page intentionally left blank

Chapter 4

Modelling Gauged Catchments – Local Procedures

Imagination is more important than knowledge, because knowledge is limited.

Albert Einstein (1879–1955)

4.1 Introduction

This introduction builds on the background discussed earlier. The underlying issues are restated here for completeness of the argument.

Many if not most rainfall-runoff model structures currently used can be classified as conceptual, as defined by Wheater *et al.* (1993). Conceptual rainfall-runoff (CRR) model structures commonly aggregate, in space and time, the hydrological processes occurring in a catchment, into a number of key responses represented by storage components (state variables) and their interactions (fluxes); see Fig. 4.1. The model parameters describe aspects such as the size of those storage components, the location of outlets, the distribution of storage volumes, etc. Conceptual parameters, therefore, usually refer to a collection of aggregated processes and may cover a large number of subprocesses that cannot be represented separately or explicitly (Van Straten and Keesman, 1991). The underlying assumption, however, is that these parameters are, even if not measurable properties, at least constants and representative of inherent properties of the natural system (Bard, 1974, p.11).

The modeller's task is the identification of an appropriate CRR model (or models) for a specific case, with a given modelling objective, catchment characteristics, and data set (Wagener, 1998). Experience shows

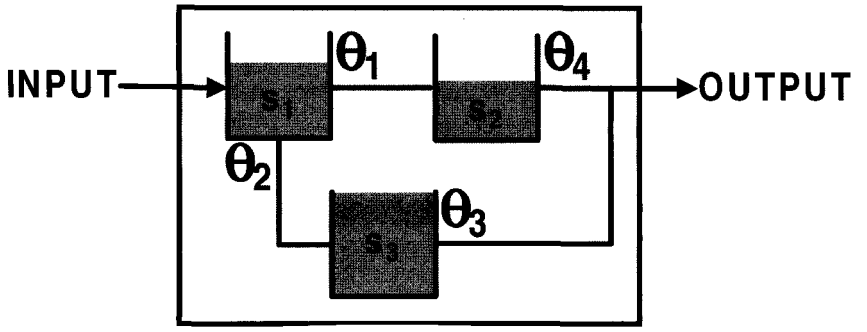


Fig. 4.1 Schematic description of a conceptual model structure showing different internal states s_i and parameters θ_i (modified from Beck, 1993).

that this identification is a difficult task. Various parameter sets, often widely distributed within the feasible parameter space (e.g., Duan *et al.*, 1992; Freer *et al.*, 1996), sometimes even different conceptualizations of the catchment system (e.g., Uhlenbrock *et al.*, 1999), may yield equally good results in terms of a predefined objective function (OF). This ambiguity has serious impacts on parameter and predictive uncertainty (e.g., Beven and Binley, 1992) and therefore limits the applicability of CRR models, for example for the simulation of land use or climate change scenarios, or for regionalization studies (Moore and Clarke, 1981; Wheater *et al.*, 1993).

Initially it was thought that this problem would disappear with improved automatic search algorithms capable of locating the global optimum on the response surface (e.g., Duan *et al.*, 1992). However, powerful global optimization algorithms are available today, while single-objective calibration procedures still cannot completely replace manual calibration. One reason for this is that the resulting hydrographs are often perceived by the hydrologist to be inferior to those produced through manual calibration (Gupta *et al.*, 1998; Boyle *et al.*, 2000). It has been suggested that this is due to the fundamental problem that single-objective automatic calibration is not sufficiently sophisticated to replicate the several performance criteria implicitly or explicitly used by the hydrologist in manual calibration. This problem is increased by indications that, due to structural inadequacies, one parameter set might not adequately describe all response modes of a hydrological system. There-

fore, there is a strong argument that the process of identification of dynamic, conceptual models has to be rethought (Gupta *et al.*, 1998; Gupta, 2001).

Three reactions to this problem of ambiguity of system description can be found in the hydrological literature. The first is the increased use of parsimonious model structures (e.g., Jakeman and Hornberger, 1993; Young *et al.*, 1996; Wagener *et al.*, 2002a), structures that only contain those parameters, and therefore model components, that can be identified from the observed system output. However, the increase in identifiability is bought at the price of a decrease in the number of processes described separately by the model. There is therefore a danger of building a model (structure) that is too simplistic for the anticipated purpose. Such a model (structure) can be unreliable outside the range of catchment conditions on which it was calibrated, such as climate and land use, due to the restriction to *justifiable* components (Kuczera and Mroczkowski, 1998). It is also particularly important that the data used does have a high information content, to ensure that the main response modes are excited during calibration (Gupta and Sorooshian, 1985; Yapo *et al.*, 1996).

The second reaction is the search for calibration methods that make better use of the information contained in the available data time-series, e.g., streamflow and/or groundwater levels. Various research efforts have shown that the amount of information retrieved using a single OF is sufficient to identify only three to five parameters (e.g., Beven, 1989; Jakeman and Hornberger, 1993; Gupta, 2001). Most CRR model structures contain a larger number. More information can become available through the definition of multiple OFs to increase the discriminative power of the calibration procedure (e.g., Gupta *et al.*, 1998; Gupta, 2001). These measures can either retrieve different types of information from a single time-series, e.g., streamflow (e.g., Wheater *et al.*, 1986; Gupta *et al.*, 1998; Dunne, 1999; Boyle *et al.*, 2000; Wagener *et al.*, 2001), or describe the performance of individual models with respect to different measured variables, e.g., groundwater levels (e.g., Kuczera and Mroczkowski, 1998; Seibert, 2000), saturated areas (Franks *et al.*, 1998), or measurements of streamflow salinity (Mroczkowski *et al.*, 1997; Kuczera and Mroczkowski, 1998). However, the usefulness of additional data can depend on the adequacy of the model structure investigated.

Lamb *et al.* (1998) found that the use of groundwater levels from one or only a few measurement points as additional output variable(s) helped to reduce the parameter uncertainty of Topmodel (Beven and Kirkby, 1979; Beven *et al.*, 1995). The use of many (>100) points, however, led to an increase in prediction uncertainty indicating structural problems in the model. Seibert and McDonnell (2002) showed in a different approach how the parameter space can be constrained when soft data (qualitative knowledge of the catchment behaviour) is included in the calibration process. The soft data in their case included information derived through experimental work about the contribution of new water to runoff and the restriction of parameter ranges to a desirable range. The result was a more realistic model, which, however, yielded sub-optimal performances with respect to many specific OFs, in this case the NSE measure.

Thirdly, some researchers abandoned the idea of a uniquely identifiable model in favour of the identification of a model population (e.g., Van Straten and Keesman, 1990; Beven and Binley, 1992; Gupta *et al.*, 1998). This can be a population of models with varying degrees of likelihood that can be representative for the catchment at hand, the idea behind the GLUE approach (Beven and Binley, 1992; Freer *et al.*, 1996). Or it may be an approach based on the recognition that the calibration of a rainfall-runoff model is inherently a multi-objective problem, resulting in a population of non-dominated parameter sets (Goldberg, 1989, p.201) in the presence of model structural inadequacies (Gupta *et al.*, 1998).

The aim of the research presented here is to increase the amount of information made available from an output time-series and to guide the identification of parsimonious model structures, consistent with a given model application as explained below. Multi-objective approaches are used for performance and identifiability analysis and a novel dynamic identifiability analysis (DYNIA) method is used for assumption testing. These can be integrated into a framework for model corroboration and rejection. Application examples at the end of this chapter show the use of these approaches and the framework for different cases.

4.2 Identification of Conceptual Rainfall-Runoff Models

The purpose of identifiability analysis in CRR modelling is to find the model structure(s) and corresponding parameter set(s) that are representative of the catchment under investigation, while considering aspects such as modelling objectives and available data. This identifiability analysis can be split into two stages: model structure selection or identification, and parameter estimation or identification. They may not, however, be treated as completely separate (Sorooshian and Gupta, 1985), because in order to evaluate model structures fully, one has to analyse their performance and behaviour, requiring some form of parameter estimation.

Traditional modelling procedures commonly contain, among others, an additional third step (e.g., Anderson and Burt, 1985; Anderson and Woessner, 1992). This is a validation or verification step often used to show that the selected model is an acceptable representation of the catchment under investigation. This results in the following three steps, which form the core of the more complete sequence developed in Chapter 2 (Fig. 2.2):

- Selection or development of a model structure and subsequent computer code to represent the conceptualization of the hydrologic system that the hydrologist has conceived for the catchment under study.
- Calibration of the selected model structure, *i.e.,* estimation of the *best* parameter set(s) with respect to one or more (often combined) criteria.
- Validation or verification of this model by successfully applying it to a data set not used during the calibration stage.

It is important to stress that the original meanings of the words 'validation' and 'verification' are different. Verification is the stronger statement, meaning *to establish the truth*, while validation means *to establish legitimacy* (Oreskes *et al.*, 1994). In the context of hydrological modelling, these terms are often applied synonymously, describing a step used to justify the chosen model as an acceptable representation of the real system. An in-depth discussion on this topic can be found in Oreskes *et al.* (1994).

These three steps are similar to the logic of *induction* often used in science. This idea of induction is founded on the underlying assumption that a general statement can be inferred from the results of observations or experiments, *i.e.,* singular statements (Popper, 2000, p.27). It includes the assumption that a hypothesis, e.g., a chosen model structure, can be shown to be correct, *i.e.,* a hypothesis can be validated or verified, through supporting evidence. The steps taken in this traditional scientific method are usually similar to the following (modified from Magee, 1977, p.56):

- observation and experiment;
- inductive generalization, *i.e.,* a new hypothesis;
- attempted verification of hypothesis, *i.e.,* proof or disproof of hypothesis;
- knowledge.

However, the logical error in this approach is, (as Magee, 1977, p.20, derives from statements by the philosopher Hume), that "no number of singular observation statements, however large, could logically entail an unrestrictedly general statement." In rainfall-runoff modelling this is equivalent to the statement that, however often a model is capable of reproducing the response of a particular catchment, it can never be concluded that the true model is found. It could for example be that future measurements will capture more extreme events, exciting a response not represented by earlier data and therefore not included in the model. Similarly, the philosopher Popper concluded that no theory or hypothesis could ever be taken as a final truth. It can only be said that it is corroborated by every observation so far, and yields better predictions than any known alternative. It will however, always remain replaceable by a better theory or may turn out to be false at a later stage (Popper, 2000, p.33).

The idea that a model can be verified (*verus*, meaning 'true' in Latin, Oreskes *et al.*, 1994) is therefore ill-founded and alternative modelling frameworks have to be found. One such alternative approach to scientific investigation in general was suggested by Popper (2000). He realized that, while no number of correctly predicted observations can lead to the conclusion that a hypothesis is correct, a single unexplained observation can lead to the falsification of the hypothesis. Hence, he replaced the

framework of verification with a framework of falsification, allowing the testing of a hypothesis.

This framework of falsification as suggested by Popper can be outlined as follows (modified from Magee, 1977, p.56):

- the initial problem or question, often resulting from the fact that an existing hypothesis has failed;
- one (or more) proposed new hypothesis(es);
- deduction of testable propositions from the new hypothesis;
- attempted falsification of the new hypothesis by testing the propositions;
- preference established between competing hypotheses.

The procedure is repeated as soon as the new hypothesis fails. It is thus possible to search for the truth, but it is not possible to know when the truth has been found, a problem which has to be reflected in any scientific method.

Additionally, Beven (2000a, p.304) points out that it is very likely, at least with the current generation of CRR models, that every model will fail to reproduce some of the behaviour of a catchment at some stage. However, even if one knows that a model is inadequate, one often has to use it due to the lack of alternatives. In many cases, the use of this inadequate model will be sufficient for the selected purpose. Or as Wilfried Trotter put it more generally: "In science the primary duty of ideas is to be useful and interesting even more than to be 'true'." (Beveridge, 1957, p.41). Mankin *et al.* (1977) come to a similar conclusion: "Let us dismiss the question, 'Have you proven that your model is valid?' with a quick NO. Then let us take up the more rewarding and far more challenging question: 'Have you proven that your model is useful for learning more about the ecosystem?'"

How this general idea of hypothesis falsification can be put into a framework for CRR modelling is described below.

4.2.1 Identification of model structures

A large number of CRR modelling structures are currently available. These differ, for example, in the degree of detail described, the manner in which processes are conceptualized, requirements for input and output

data, and possible spatial and temporal resolution. Despite these differences, a number of model structures may appear equally likely for a specific study, and the selection process often amounts to a subjective decision by the modeller (Wagener, 1998), since objective decision criteria are often lacking (Mroczkowski *et al.*, 1997). It is therefore important to deduce *testable propositions* with respect to the assumptions underlying the model structure, *i.e.,* on how the catchment works, and to find measures of evaluation that give some objective guidance as to whether a selected structure is suitable or not. Uhlenbrock *et al.* (1999) have shown however that it is difficult to achieve this using single-objective Monte Carlo-based calibration approaches. They were able to derive good performances with respect to the prediction of streamflow, from both sensible, as well as incorrect conceptualizations of a catchment. Mroczkowski *et al.* (1997) encountered similar problems when trying to falsify one of two possible model structures, including and excluding a groundwater discharge zone, to represent two paired catchments in Western Australia. This was impossible for both catchments when only streamflow data were used. The additional use of stream chloride and groundwater level measurements allowed at least for the falsification of one of the model structures for the second catchment which had undergone considerable land use changes.

Testable propositions about a specific model structure can either be related to the *performance* of the model or its components, or they can be related to its proper *functioning*.

A test of performance is the assessment of whether or not the model structure is capable of sufficiently reproducing the observed behaviour of the natural system, considering the given quality of data. However, an overall measure of performance, aggregating the residuals over the calibration period and therefore a number of response modes, hides information about how well different model components perform. It can be shown that the use of multiple-objectives for single-output models, measuring the model's performance during different response modes, can give more detailed information and can allow the modeller to link model performance to individual model components (e.g., Boyle *et al.*, 2001; Wagener *et al.*, 2001). Additional information will also be available in cases where the model produces other measurable output vari-

ables, e.g., groundwater levels or hydro-chemical variables, as mentioned earlier.

Evaluation of the proper functioning of the model means questioning the assumptions underlying the model's structure. Do the model components really represent the response modes they are intended to represent? Is the model structure capable of reproducing the different dominant modes of behaviour of the catchment with a single parameter set? A model structure is usually a combination of different hypotheses of the working of the natural system. If those hypotheses are to be individually testable, they should be related to individual model components and not just to the model structure as a whole (Beck, 1987; Beck *et al.*, 1993).

One previously mentioned, underlying assumption of conceptual modelling is the consideration of model parameters as constant in time, at least as long as no changes in the catchment occur that would alter the hydrological response, such as land use changes. Different researchers (e.g., Beck, 1985; 1987; Gupta *et al.*, 1998; Boyle *et al.*, 2000; Wagener *et al.*, 2003a, 2000c) have shown that this assumption can be tested, and that the failure of a model structure to simulate different response modes with a single parameter set suggests inadequacies in the functioning of the model.

Beck used the Extended Kalman Filter (EKF) extensively to recursively estimate model parameters and to utilize the occurrence of parameter deviation as an indicator of model structural failure (e.g., Beck, 1985; 1987; Stigter *et al.*, 1997). For example, in the identification of a model of organic waste degradation in a river, changes were identified in optimum parameter values in time from one location in the parameter space to another (Beck, 1985). Beck showed that the model hypothesis had failed; the parameters were changing to compensate for one or more missing aspects in the model structure. The subsequent step is to draw inference from the type of failure to develop an improved hypothesis of the model structure. However, there are limitations to the EKF approach. Beck (1987) concluded with respect to its use for hypothesis testing that "the performance of the EKF is not as robust as would be desirable and, inter alia, is heavily compromised by the need to make more or less arbitrary assumptions about the sources of uncertainty affecting the identification problem."

A trade-off in the capability to simulate different response modes can occur, as shown by Boyle *et al.* (2000) for the example of a popular complex rainfall-runoff model (Sacramento with 13 calibrated parameters; Burnash, 1995): thus, they found it was not possible to reproduce slow recession periods and the remaining system response modes simultaneously. Their multi-objective analysis suggests the cause for this problem is mainly inadequate representation of the upper soil zone processes.

Approaches to address the problem of model structure identification in a more objective way are therefore available. However, they are not without weaknesses, as Beck's statement about the use of the EKF showed. These weaknesses need to be addressed to derive approaches more suitable for the nature of CRR model structures.

4.2.2 Identification of parameters

The second stage in model identification is the estimation of a suitable parameter set, usually referred to as calibration of the model structure. In this process, the parameters of a model structure are adjusted until the observed system output and the model output show acceptable levels of agreement. Manual calibration does this in a trial-and-error procedure, often using a variety of measures of performance and visual inspection of the hydrograph (e.g., Gupta *et al.*, 1998). It can yield good results, but is time consuming, requires extensive experience with a specific model structure, and precludes an objective analysis of parameter uncertainty. Traditional single-criterion automatic calibration on the other hand is fast and objective, but will produce results that reflect the choice of OF and may therefore not be acceptable to hydrologists concerned with a number of aspects of performance (Boyle *et al.*, 2000; Hogue *et al.*, 2000). In particular, the aggregation of model residuals into an OF leads to the neglect and loss of information about individual response modes, and can result in a biased performance, fitting a specific aspect of the hydrograph at the expense of another. It also leads to problems with the identification of those parameters associated with response modes that do not significantly influence the chosen OF (Wagener *et al.*, 2001). For example, selecting an OF that puts more emphasis on fitting peak flows, e.g., the NSE (Nash and Sutcliffe, 1970), due to its use of squared residual values

(Legates and McCabe, 1999), often will not allow for the identification of parameters related to the slow response of a catchment (e.g., Dunne, 1999; Hogue *et al.*, 2000).

An example to demonstrate this problem is briefly presented. It uses a simple model structure consisting of a Penman two-layer soil moisture accounting component (Penman, 1949) to produce effective rainfall, and a linear routing component using two conceptual reservoirs in parallel to transform it into streamflow (see Chapter 3 for details). A comparison of hydrographs produced by different parameter sets within the selected structure, which yield similar OF values, shows that these hydrographs can be visually different. Figure 4.2 shows a hundred days extract from six years of daily streamflow data, where the observed time-series (black line) is plotted with seven different realizations (grey lines), *i.e.,* the same model structure is used with different parameter sets. The OF used during calibration is the root mean squared error (RMSE), which can be defined as follows:

$$RMSE(\theta) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(y_k - \hat{y}_k(\theta)\right)^2} \tag{4.1}$$

where $\hat{y}_k(\theta)$ is the calculated flow at time step k using parameter set θ, and y_k is the corresponding observed flow, while N is the total number of time steps considered. Each of the models presented yields an RMSE of 0.60 mmd^{-1} when the complete calibration period (6 years) is considered. However, the hydrographs produced are clearly visually different. The added dotty plots of the two residence times of the (linear) routing component show that while the quick flow residence time, k(quick) is very well identified, the slow flow residence time, k(slow), is not. This is consistent with the observation that the main difference between the hydrographs can be observed during low flow periods. The effect is due to the use of squared residuals when calculating the RMSE.

This demonstrates that traditional single-objective optimization methods often do not have the ability to distinguish between *visually different behaviour* (Gupta, 2001). The requirement for a parameter set to be uniquely locatable within the parameter space, *i.e.,* to be globally identifiable, is that it yields a unique response vector (Kleissen *et al.*,

1990; Mous, 1993). The unique response vector, in this case a calculated hydrograph, might be achievable, but this uniqueness may be lost if the residuals are aggregated into a single OF. Such problems cannot be resolved through improved search algorithms. They are rather inherent in the philosophy of the calibration procedure itself.

Clearly, the complex thought processes that lead to successful manual calibration are very difficult to encapsulate in a single OF. This is illustrated by the requirements defined by the US National Weather Service (NWS) for the manual calibration of the Sacramento model structure (NWS, 2001):

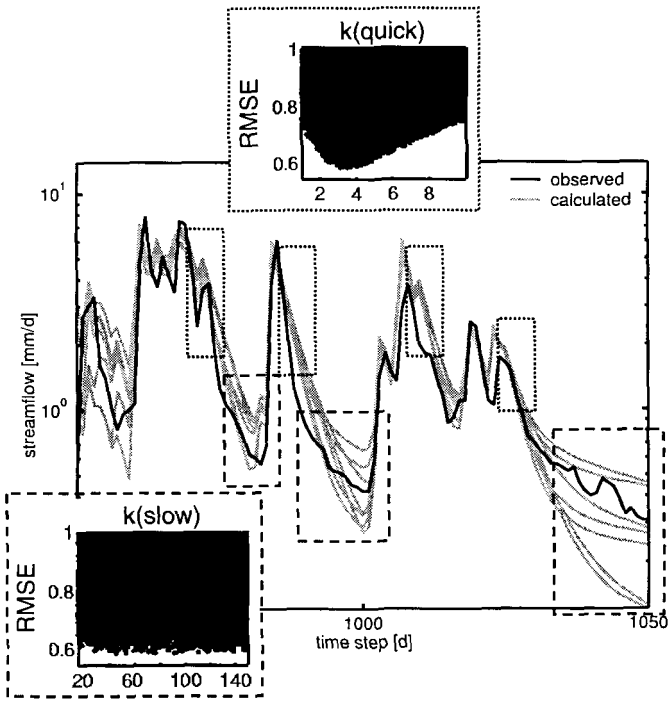


Fig. 4.2 Hundred days' extract of six years of daily streamflow data. Observed flow is in black, seven different model realizations are in grey. Inserts show dotty plots for the time constants k(quick) and k(slow) with respect to the Root Mean Squared Error (RMSE). The model structure used consists of a Penman soil moisture accounting component and a parallel routing component of linear reservoirs with fixed flow distribution (see application example 2 for details). (From Wagener *et al.*, 2003b; reproduced with permission of the American Geophysical Union)

- Proper calibration of a conceptual model should result in parameters that cause model components to mimic processes they are designed to represent. This requires the ability to isolate the effects of each parameter.
- Each parameter is designed to represent a specific portion of the hydrograph under certain moisture conditions.
- Calibration should concentrate on having each parameter serve its primary function rather than overall goodness of fit.

It can be seen from these requirements that manual calibration is more complex than the optimization of a single OF, and that traditional single-criterion automatic calibration procedures will in general not achieve comparable results. It is for example often not possible to isolate the effects of individual parameters and treat them as independent entities as done in the manual approach described above. Another aspect is that the goal of single-objective optimization is purely to optimize the model's performance with respect to a selected overall goodness of fit measure which is very different from the third requirement shown above. This is not to say that traditional individual OFs are not important parts of any model evaluation. The point is rather that they are not sufficient and should be complemented by a variety of measures.

Gupta *et al.* (1998) review this problem in more detail and conclude that a multi-objective approach to automatic calibration can be successful. Boyle *et al.* (2000) show how such a procedure can be applied to combine the requirements of manual calibration with the advantages of automatic calibration. A multi-objective algorithm is used to find the model population necessary to fit all aspects of the hydrograph. The user can then, if necessary, manually select a parameter set from this population to fit the hydrograph in the desired way. However, in the presence of model structural inadequacies this will lead to a sub-optimal performance with respect to at least some of the other measures (Boyle *et al.*, 2000; Seibert and McDonnell, 2002). The resulting trade-off in the ability of different parameter sets to fit different aspects of the hydrograph usually leads to a *compromise solution* (Ehrgott, 2000) in cases where a single parameter set has to be specified. The procedure of Boyle *et al.* (2000) for example, analyses the *local behaviour* of the model additionally to its *global behaviour* (Gupta, 2000). The global behaviour is described

through OFs such as overall bias or some measure of the overall variance, e.g., the RMSE. The local behaviour is defined by aspects such as the timing of the peaks, or performance during quick and slow response periods (Boyle *et al.*, 2000; 2001).

Recent research into parameter identification has thus moved away from simply trying to improve search algorithms to taking a closer look at the assumptions underlying (automatic) calibration approaches (e.g., Gupta *et al.*, 1998). This has lead to the use of multi-objective (MO) automatic techniques which so far have given promising results (Boyle *et al.*, 2000; Wagener *et al.*, 2001). Further investigations are required to make MO optimization a standard method for parameter estimation. Questions such as the appropriate number and derivation of OFs within an MO approach must be resolved, and will probably depend on model structure and catchment characteristics (Gupta, 2001). A method of finding the appropriate model population while considering, for example, the uncertainty in the individual parameter estimates has been defined by Vrugt *et al.* (2003b).

4.3 Evaluation of Conceptual Rainfall-Runoff Models

It was established earlier that the idea of calibration and validation of CRR models is in principle ill-founded if the goal is to establish a model as the true representation of a hydrological system. The model identification problem is therefore seen here as a process of model evaluation. Within this process, models and model structures are evaluated with respect to different criteria and those that fail, in whatever way, are rejected as possible representations of the catchment under investigation. This usually results in a population of feasible models or even model structures that can then be used for a combined prediction, resulting in a prediction range rather than a single value for each time-step. This evaluation should be done in at least in three dimensions:

- *performance*, with respect to reproducing the behaviour of the system;
- *uncertainty* in the parameters, which is assumed to be inversely related to their identifiability; and

- *assumptions*, *i.e.*, are any assumptions made during the development of the model structure violated?

The smaller the population of models (or even model structures) that survives this evaluation (*i.e.*, those that are corroborated by it), the better identifiable is the representation of the natural system in mathematical form. Approaches to test models with respect to these three criteria are described below.

4.3.1 Evaluation of competing model structures – Multi-objective performance and identifiability analysis

Multi-objective (MO) approaches can be applied to establish preferences between competing model structures or even model components, (competing hypotheses), with respect to their performance and identifiability. An MO approach is advantageous because "the use of multiple objective criteria for parameter estimation permits more of the information contained in the data set to be used and distributes the importance of the parameter estimates among more components of the model. Additionally, the precision of some parameters may be greatly improved without an adverse impact on other parameters." (Yan and Haan, 1991).

4.3.1.1 Measures of performance and identifiability

It has been shown that it is advantageous to evaluate the global and the local behaviour of models to increase the amount of information retrieved from the residuals in the context of single output RR models. Global behaviour is measured by traditional OFs (e.g., the RMSE or the bias for the whole calibration period), while different OFs must be defined to measure local behaviour. One way of implementing local measures is by partitioning the continuous output time series into different response periods. A separate OF can then be specified for each period, reducing the amount of information lost through aggregation of the residuals, for example by mixing high flow and recession periods.

Partitioning schemes proposed for hydrological time series include those based on (a) *Experience* with a specific model structure (e.g., the Birkenes model structure in the case of Wheater *et al.*, 1986): different

periods of the streamflow time series are selected based on the modeller's judgement. The intention of Wheater *et al.* (1986) was to improve the identifiability of insensitive parameters (so-called minor parameters), with respect to an overall measure. Individual parameters or pairs of parameters are estimated using a simple grid search to find the best values for individual OFs. This is done in an iterative and sequential fashion, beginning with the minor parameters and finishing with the dominant ones. (b) *Hydrological understanding*, the separation of different catchment response modes through a segmentation procedure based on the hydrologist's perception of the hydrological system (e.g., Harlin, 1991; Dunne, 1999; Boyle *et al.*, 2000; Wagener *et al.*, 2001). For example, Boyle *et al.* (2000) propose hydrograph segmentation into periods *driven* by rainfall and periods of drainage. The drainage period is further subdivided into quick and slow drainage by a simple threshold value. (c) *Parameter sensitivity* (e.g., Kleissen, 1990; Wagner and Harvey, 1997; Harvey and Wagner, 2000), where it is assumed that informative periods are those time-steps during which the model output shows a high sensitivity to changes in the model parameters (Wagner and Harvey, 1997). Kleissen (1990) developed an optimization procedure whereby only data segments in which the parameter shows a high degree of first order sensitivity are included in the calibration of that parameter (group), utilising a local optimization algorithm. (d) *Similar characteristics* in the data, derived from techniques like cluster analysis (e.g., Boogard *et al.*, 1998) or wavelet analysis (Gupta, 2001) can be used to group data points or periods based on their information content. The different clusters could then be used to define separate OFs.

While these methods help to retrieve more information, they also show some weaknesses. Approaches (a) and (b) are subjectively based on the hydrologist's experience, and therefore not easily applicable to a wide variety of models and catchments. Approach (c), while objective, does not recognize the effects of parameter dependencies, and may not highlight periods that are most informative about the parameters as independent entities, *i.e.*, periods where the dependency with respect to other parameters is low. The sensitivity of the model performance to changes in the parameter is a necessary requirement, but it is not sufficient for the identifiability of the parameter. Furthermore, if the parameter sensitivity

is measured locally (e.g., Kleissen, 1990), the result is not guaranteed over the feasible parameter space. However, Wagner and Harvey (1997) show this problem can be reduced by implementing a Monte Carlo procedure where sensitivity for a large number of different parameter combinations is assessed using parameter covariance matrices. Approach (d) is independent of any model structure and links between the results and the model parameters still need to be established.

There is therefore room to improve the objectivity, applicability and robustness of approaches to hydrograph disaggregation, with the goal of improving model structure and parameter identifiability.

The evaluation of the model performance should, if possible, also include OFs tailored to fit the specific purpose of the model. An example is the use of the model to investigate available quantities for abstraction purposes. Assuming that abstraction can only take place during periods when the water level is above a minimum environmentally acceptable flow and below a maximum water supply abstraction rate allows the definition of a specific OF (Lees and Wagener, 2000). This measure can aggregate the residuals of the selected period and provide important information about how a model performs with respect to the anticipated task. However, it is important to mention that this should never be the sole evaluation criterion.

How, then, can one estimate the identifiability of the individual parameters with respect to the different OFs defined? Wagener *et al.* (2001; see also Section 3.3.3) define a simple empirical measure of identifiability. It is based on the parameter population conditioned by a selected measure of performance (Fig. 4.3). A uniform random sampling procedure is performed, and resulting OF values are transformed so that the best performing parameter set is assigned the highest value and all measures sum to unity (these are termed performance values in Fig. 4.3a). The best performing 10% of all parameter sets are selected and the cumulative marginal distributions for each parameter are plotted (Fig. 4.3b). A uniform distribution would plot as a straight line, while a distribution showing a clear peak will show a steeper line. The stronger the conditioning, the larger the steepness will be. The range of each parameter is subsequently split into M containers and the gradient of the cumulative distribution in each container is calculated (Fig. 4.3c). The highest gradi-

ent occurs where the conditioning of the distribution is strongest, *i.e.*, at the location of a peak. The amplitude of the gradient is also indicated by the grey shading of the bar (Fig. 4.3b), with a darker colour indicating a higher gradient. Other measures of identifiability are possible (e.g., Wagener *et al.*, 1999; Appendix A), but this one has proven to be robust and easy to calculate.

4.3.1.2 Multi-objective framework

The above-described multi-objective performance and identifiability approaches can be put into an analytical framework to estimate the appropriate level of model complexity for a specific case (Fig. 4.4, adapted from Wagener *et al.*, 2001).

The hydrologist's perception of a given hydrological system strongly influences the level of conceptualization that must be translated into the model structure. The importance of different system response modes, *i.e.*, key processes that need to be simulated by the model, however, depends on the intended modelling *purpose*. Therefore, the level of model structural complexity required must be determined through careful consideration of the key processes included in the model structure and the level of prediction accuracy necessary for the intended modelling purpose.

On the other hand there is the level of structural complexity actually supported by the information contained within the observed data. It is defined here simply as the number of parameters, and therefore separate model components (and possibly processes) that can be identified. Other aspects of complexity (e.g., Kleissen *et al.*, 1990) such as the number of model states or interactions between the state variables or the use of non-linear components instead of linear ones, are not considered here.

An increase in model complexity will often improve the performance. However, it will also often increase the uncertainty, for example due to reduction in parameter identifiability caused by enhanced parameter interaction. An example of this trade-off can be seen in Fig. 4.5, where different model structures are compared with respect to performance and identifiability using data from the Medway at Teston (see Section 4.4.2.1 for details). Four OFs, based on the RMSE and describing different parts of the hydrograph, are used (see Section 4.4.3). Calculating the arithme-

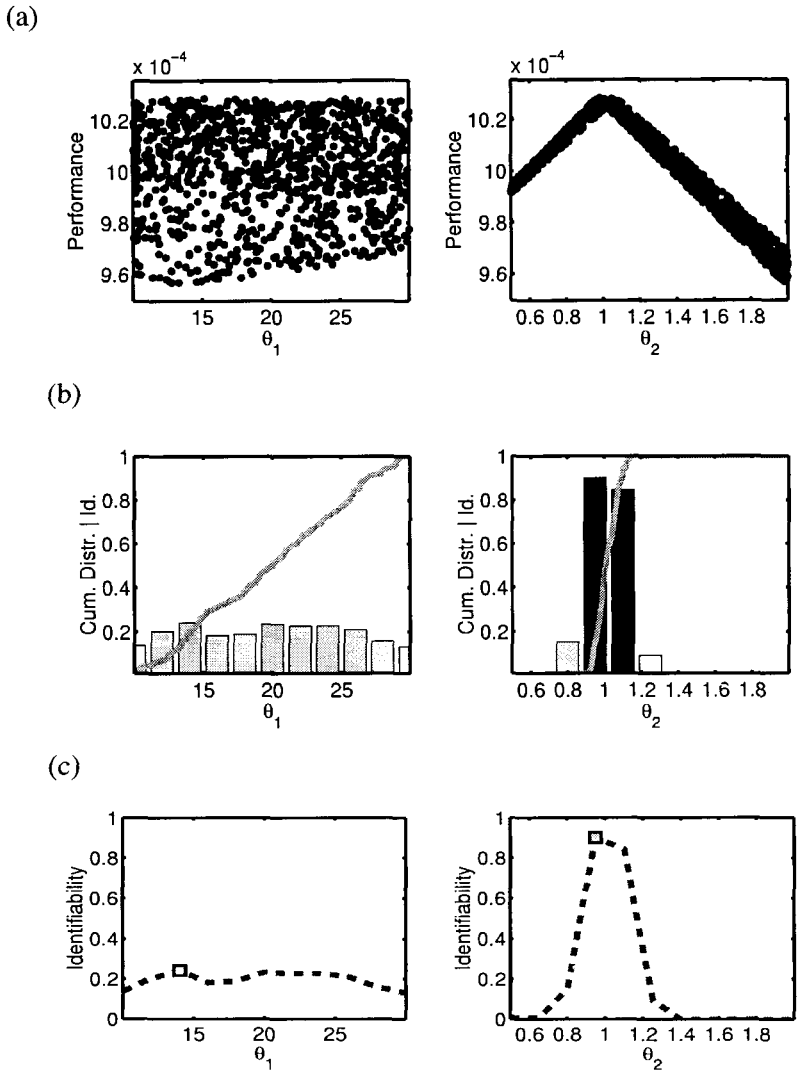


Fig. 4.3 A measure of identifiability can be defined as follows: an initially uniform distribution is conditioned on some OF, the resulting dotty plot is shown in plots (a); the top percentile (e.g., 10%) is selected and plotting the cumulative distribution of the transformed OF values leads to plots (b); the gradient distribution of the cumulative distribution is a measure of identifiability (c). The plots in the right column show an identifiable parameter, while the plots in the left column show a non-identifiable one. (From Wagener *et al.*, 2001; reproduced with permission of the European Geophysical Society)

tic mean of all OFs yields a single aggregated measure of performance for individual models and a single aggregated measure of identifiability for individual parameters. The mean of the individual parameter identifiability values gives a value for the model as a whole. It can be seen that some improvement in the optimum direction, *i.e.,* the bottom left corner, is possible. However, unavoidably, the best performing model will general not be the most identifiable one and vice versa. What trade-off, between performance and identifiability is acceptable depends on the modelling purpose and the hydrologist's preference. In a regionalization study, a more identifiable model with reduced performance might be adequate, while parameter identifiability might be of less importance for extension of a single-site record. It must be stressed that Fig. 4.5 is only qualitative; the location of points will vary with the aggregation scheme used, and a high performance does not mean that a visually acceptable hydrograph is produced.

It has already been established (Section 4.2.2) that such a framework has to use a multi-objective approach to allow for objective analysis. Using various OFs to represent different system response modes is especially suitable for comparison studies, since it allows us to attribute the model performance during different system response modes to different model components, such as moisture accounting or routing (Wagener *et al.*, 2001). Using the segmentation approach by Boyle *et al.* (2000) as described earlier (Section 3.4), it is possible to establish that a specific model structure might perform better during *driven* periods because of a superior moisture accounting component, while another model structure containing a more appropriate slow flow routing component could result in higher performance during *non-driven slow* periods. A single-objective framework does not allow comparison of model components and consequently important information relevant to identifying the most suitable model structure is lost. Boyle *et al.* (2001) use the segmentation approach to evaluate the benefit of *spatial distribution* of model input (precipitation), structural components (soil moisture and streamflow routing computations), and surface characteristics (parameters) with respect to the reproduction of different response modes of the catchment system.

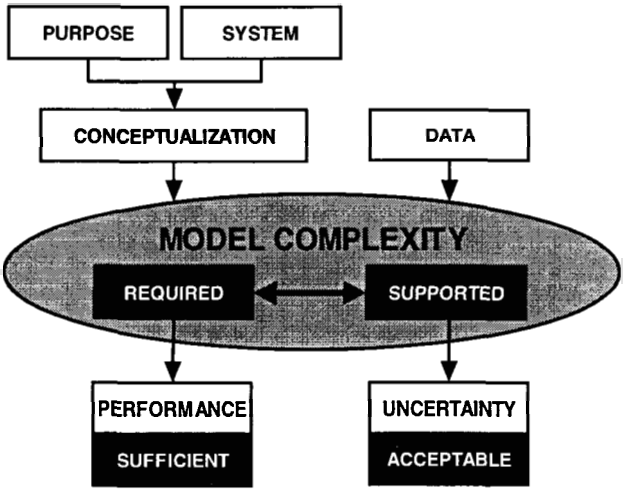


Fig. 4.4 Framework for the evaluation of competing rainfall-runoff model structures. (From Wagener *et al.*, 2003b; reproduced with permission of the American Geophysical Union)

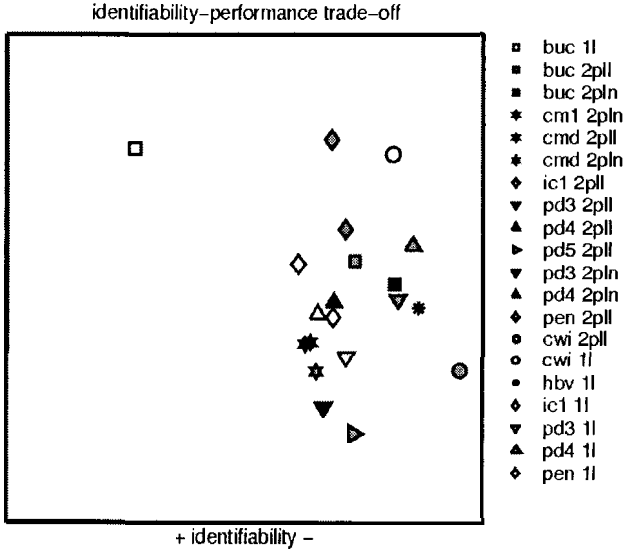


Fig. 4.5 Scatter plot showing an identifiability-performance trade-off. The abbreviations can be found in the Notation Section.

This framework will also necessarily be comparative; different models and usually different model structures will have to be compared to identify one or more suitable models. The reason is that the level of performance that can be reached is unknown, due to unknown influences of data error and of natural randomness. Models and model structures that severely under-perform can be refuted and excluded from further consideration. In cases where all models fail, one must go back and relax the criteria for under-performance (Beven, 2000a, p.304).

Model structures that produce more than a single output variable (e.g., groundwater levels or water quality) can be tested with respect to all of those variables if measurements are available (Gupta *et al.*, 1999; Bastidas *et al.*, 1999). One could say that the *informative* (or *empirical*) *content* of these structures is higher and they have, therefore, a higher degree of *testability* or *falsifiability* (Popper, 2000, p.113). However, a hypothesis (or a model structure, in our case) that has a higher information content is also logically less probable, because the more information a hypothesis contains, the more options there are for it to be false (Popper, 2000, p.119; Magee, 1977, p.36). Multi-output models are beyond the scope of this monograph though.

4.3.2 Evaluation of individual model structures — Dynamic identifiability analysis

There is an apparent lack of objective procedures to evaluate the suitability of an individual conceptual model structure to represent a specific hydrological system. It has been shown how different and competing structures can be compared. However, it is also possible to analyse individual structures with respect to the third criterion mentioned in Section 4.3, model assumptions.

4.3.2.1 Failure, inference and improved hypotheses

Recently, Gupta *et al.* (1998; see also Boyle *et al.*, 2000 and Wagener *et al.*, 2001) showed how a multi-objective approach can be applied to give an indication of structural inadequacies. The assumption is that a model should be capable of representing all response modes of a hydrological

system with a single parameter set. A failure to do so indicates that a specific model hypothesis is not suitable and should be rejected or replaced by a new hypothesis that improves on the old one. This idea has a basis in some of Beck's work (e.g., Beck, 1985, 1987) as described earlier in the text. A new approach has been developed based on this assumption, as reported by Wagener *et al.* (2003a) and presented in the next section. This methodology analyses the identifiability of parameters within a selected model structure in a dynamic and objective manner (considering the assumptions made), which can be used to analyse the consistency of locations of those parameter values that give a good performance in (parameter) space and in time.

In cases where the variation of parameter optima can be tracked in time it is sometimes possible to directly relate changes in a particular parameter to variations in forcing or state variables (examples are shown in Beven, 2000a, p.93ff.; Bashford *et al.*, 2002). However, in many cases, the development of improved hypotheses will be more complex and depend on the capability of the hydrologist. Unfortunately (or maybe fortunately), there is no logical way to create new ideas; the hydrologist therefore has to apply his depth of insight and creative imagination to derive a new hypothesis to replace the old one that has failed.

4.3.2.2 Dynamic identifiability analysis

The dynamic identifiability analysis (DYNIA) is a new approach to locating periods of high identifiably for individual parameters and to detect failures of model structures in an objective manner. The proposed methodology draws from elements of RSA (Spear and Hornberger, 1980; Hornberger and Spear, 1981) and includes aspects of the GLUE approach (Beven and Binley, 1992; Beven, 1998), wavelet analysis (e.g., Gershenfeld, 1999) and the use of Kalman filtering for hypothesis testing as applied by Beck (1985, 1993).

In the original RSA approach, a model population is sampled from a uniform distribution. This population is divided into *behavioural* and *non-behavioural* models depending on whether the model produced a certain response or not (Spear and Hornberger, 1980). Beven and Binley (1992) extended the approach by conditioning the model population on a

likelihood measure, which, in their case could be a transformation of any measure of performance. These are the building blocks from which a new method of assessing the identifiability of parameters is created (Wagener *et al.*, 2003a).

The steps taken in the procedure can be seen in the flow chart in Fig. 4.6. Monte Carlo sampling based on a uniform prior distribution is used to examine the feasible parameter space. The OF associated with each parameter set, *i.e.,* each model is transformed into a support measure; all support measures have the characteristic that they sum to unity and higher values indicate better performing parameter values. These are shown here in the form of a dotty plot (Fig. 4.6[a]). The identifiability measure introduced in Fig. 4.3 is calculated for the best performing parameter values (*e.g.,* the top 10%). Segmenting the range of each parameter (e.g., into 20 containers) and calculating the identifiability measure (gradient) in each container leads to the (schematic) distribution shown in Fig. 4.6(d). The highest value, indicated by the darkest colour, marks the location (within the chosen resolution) of greatest identifiability of the parameter. This measure of identifiability can be used to compare different model structures in terms of parameter uncertainty, which can be assumed to be inversely related to identifiability as shown by Wagener *et al.* (2001). They calculate the identifiability as a function of measures of performance for the whole calibration period and for specific response modes, derived using the segmentation approach of Boyle *et al.* (2000) described earlier. It can be shown that the identifiability of some parameters, and therefore individual model components, is greatly enhanced by this segmentation (Wagener *et al.*, 2001; see Section 4.4.1).

Calculating parameter identifiability at every time step t using only the residuals for a number of time steps n before and after the point considered, *i.e.,* using a *moving window* or *running mean* approach, allows the investigation of the identifiability as a function of time (Fig. 4.6[e]). The gradient distribution plotted at time step t therefore aggregates the residuals between $t - n$ and $t + n$, with the window size being $2n + 1$. The number of time steps considered depends upon the length of the period over which the parameter is influential. For example, investigation of a slow response linear store residence time parameter requires a wider moving window than the analysis of a quick response residence time

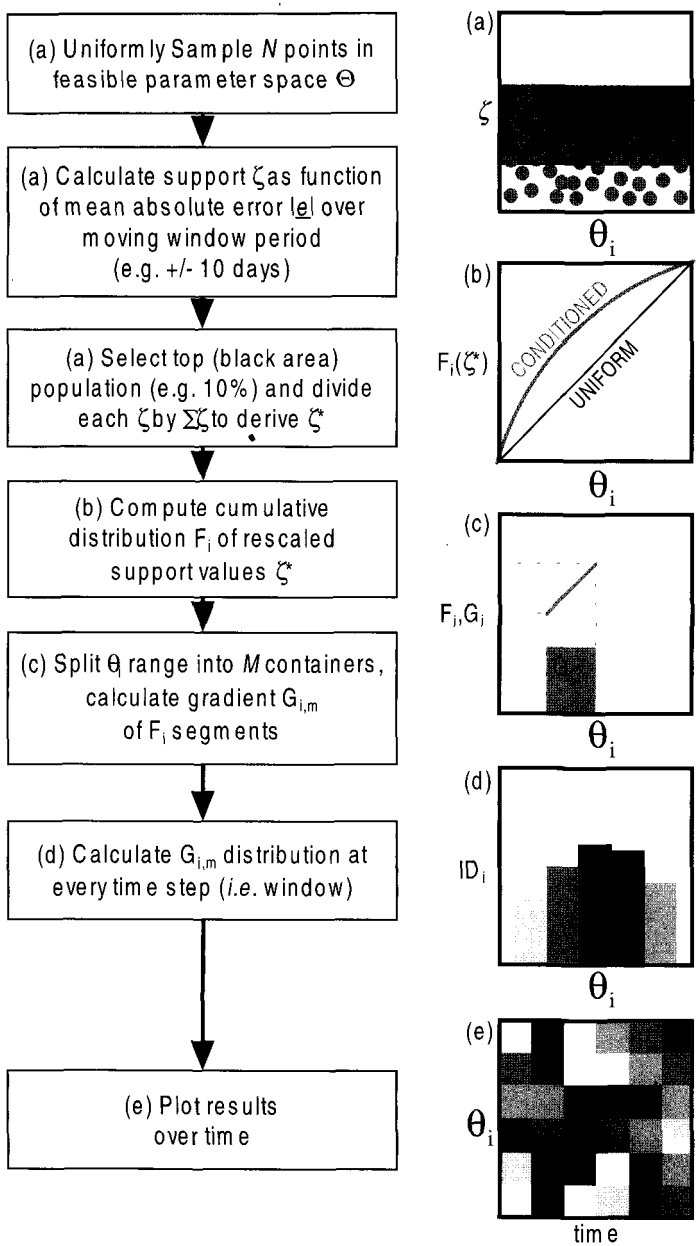


Fig. 4.6 Schematic description of the DYNamic Identifiability Analysis (DYNIA) procedure.

parameter. Different window sizes are commonly tested and the ones found to be most appropriate are used to analyse individual parameters. A window size that is too small can lead to the result being largely influenced by errors in the data. However, this is not a problem in cases where the data quality is very high, as in the case of in-stream tracer experiments (Wagener *et al.*, 2002b). Conversely, if the window size is too big, periods of noise and periods of information will be mixed and the information will be blurred.

Results are plotted for each parameter versus time using a colour coding where a darker colour indicates areas of higher identifiability in parameter space and time. Care has to be taken when interpreting the DYNIA results of time steps at the beginning and the end of time-series. Here the full window size cannot be established and the result is distorted. This is an effect similar to the *cone of influence* in wavelet analysis (Torrence and Compo, 1998).

While this approach is not intended to evaluate parameter dependencies in detail, the significance of dependencies for the identifiability is implicit in the univariate marginal distribution, which is structurally represented by Fig. 4.6(d). A strong dependency during any period would tend to inhibit the information of a strong univariate peak; the effect of the involved parameters cannot be singled out. Parameter interdependence can be estimated in detail by the investigation of the response surface or the variance-covariance matrix (e.g., Wheater *et al.*, 1986; Hornberger *et al.*, 1985).

A limitation of the proposed measure of identifiability arises if any near-optimal parameter values are remote from the identified peak of the marginal distribution, as the relevance of such values would be diminished. It is therefore important that a detailed investigation of the dotty plots be undertaken to verify periods of high identifiability.

DYNIA requires that sensible feasible ranges for each parameter be defined and that the number of models (*i.e.*, parameter sets) sampled is sufficient to represent the shape of the response surface. The procedure can then be applied to separate periods that do and do not contain information about specific parameters, and parameter variations can be tracked in time. Both aspects are shown schematically in Fig. 4.7.

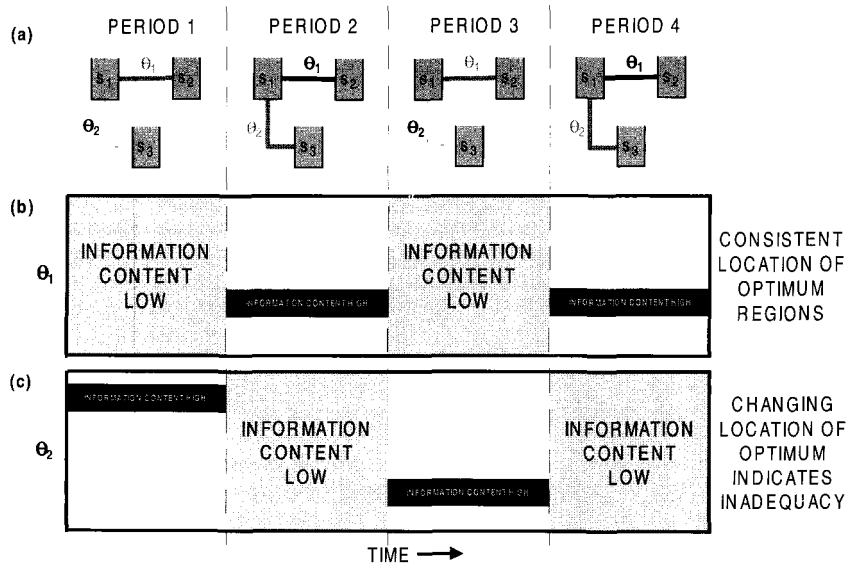


Fig. 4.7 Possible application of the DYNIA approach. Plot (a) shows a model structure with parameters θ_1 and θ_2, and the states S_1 and S_2. The structure is shown during four different response modes. A parameter in plot (a) is shown in black during periods of identifiability, and in grey during periods where it is less identifiable. It is shown as a dashed line when DYNIA indicates a failure of the component described by the parameter. Plots (b) and (c) show (schematic) DYNIA results for the parameters θ_1 and θ_2 respectively, *i.e.*, parameter identifiability versus time. A grey area indicates that equally good parameter values are widely distributed over the feasible range. A small black area means that good values are clustered in a specific region of the 1-dimensional parameter space.

4.3.3 A combined framework of corroboration and rejection

The multi-objective framework (Wagener *et al.*, 2001) introduced earlier can be extended to incorporate the DYNIA approach as an additional step deriving a framework of corroboration and rejection (Fig. 4.8). Similar frameworks are for example proposed by Beven (2000a, p.297ff.; 2002), and, more generally, by Oreskes *et al.* (1994).

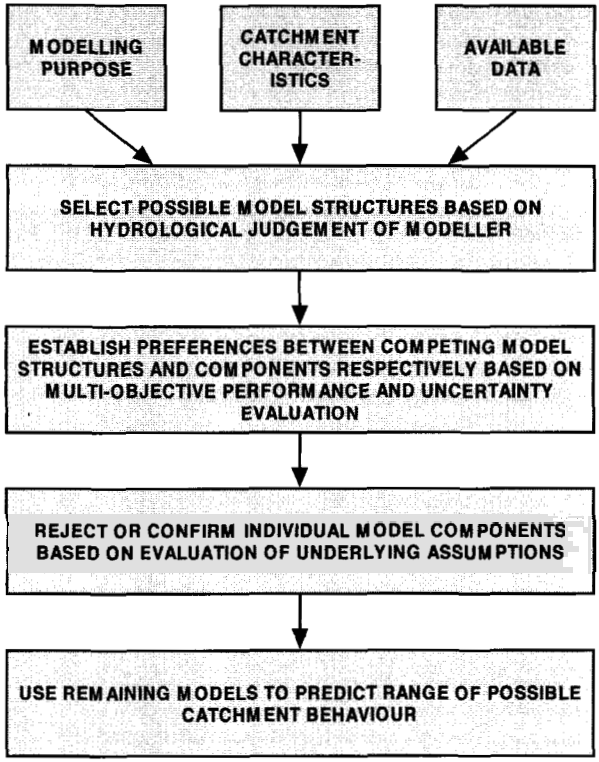


Fig. 4.8 The proposed modelling framework. (From Wagener *et al.*, 2003b; reproduced with permission of the American Geophysical Union)

The initial steps are similar to those in the multi-objective framework described earlier. The hydrologist selects or develops model structures that seem suitable for the given modelling purpose, catchment characteristics and data.

One can then apply a multi-objective procedure to establish preferences between the competing model structures, or structural components. Under-performing structures or components can be rejected at this stage, based on their performance and/or uncertainty.

During the next stage, the DYNIA approach can be used to further analyse the remaining model structures. Additional rejections might be possible. The suitability of a model structure passing this stage is further corroborated. A model structure is temporarily accepted when no better

performing structure can be found and no underlying assumption is violated.

In the last stage, the parameter space *within* the remaining model structures can be analysed to find all those models, *i.e.,* parameter sets, that are in line with the behaviour of the natural system. It is very likely that such a procedure will result in a range of acceptable or *behavioural* models or even model structures. The appropriate response is to consider the predictions of all models to derive an ensemble prediction of the system's behaviour. A popular approach to do so is the GLUE approach (Beven and Binley, 1992); however, other methods to consider or combine the predictions of different models are possible (e.g., Shamseldin *et al.*, 1997). Within the GLUE approach, a likelihood value is derived for every model. The models are usually drawn from a uniform distribution. Basically any measure of performance that can be transformed so that higher values indicate better models and all measures add up to one, can be used as a likelihood measure in this approach. The likelihoods are then used to weight the prediction of every model at every time step. The cumulative distribution of the weighted streamflow values, even for different models, allows the extraction of percentiles to derive the confidence limits for the predictions. Beven (1998) and Wagener *et al.* (2000) discuss various approaches to define a behavioural model population based on multiple criteria. The likelihoods of different model structures can, for example, be combined through simple addition. This approach is shown schematically in Fig. 4.9.

4.4 Application Examples

4.4.1 Example 1 – Multi-objective performance and identifiability analysis

A simple example is used to demonstrate how the proposed multi-objective framework can be applied. The data and model structure selected for the case study are described briefly, and examples of possible applications of the tools for model calibration and evaluation are shown.

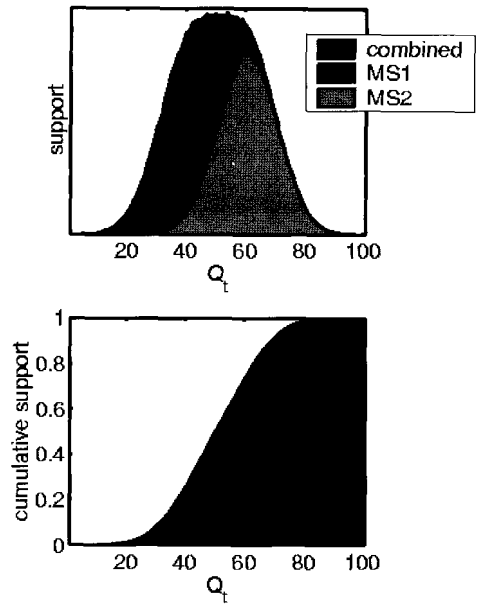


Fig. 4.9 Possible form of ensemble prediction with different model structures which could be implemented using for example the GLUE approach.

4.4.1.1 Data

The Leaf River catchment (1950 km^2) located north of Collins, Mississippi, USA, which has been extensively investigated (e.g., Brazil and Hudlow, 1981; Sorooshian *et al.*, 1983), is selected for this study. Forty consecutive water-years (1948–1988) of data (daily precipitation, streamflow, and potential evapotranspiration estimates) are available for this catchment, representing a wide variety of hydrologic conditions. An 11-year period (1952–1962 inclusive) is used here.

4.4.1.2 Model structure and calibration tools

A simple model structure with typical conceptual components is chosen for an example application of the framework. This model consists of a simple two-parameter SMA component connected with two series of linear reservoirs (three [identical] for the quick response and a single reser-

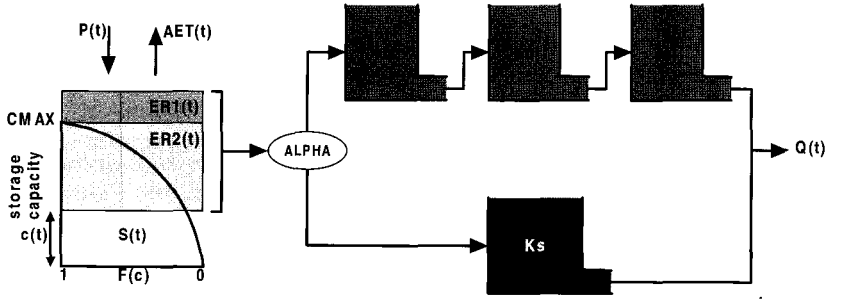


Fig. 4.10 The model structure used in the rainfall-runoff modelling example. Effective rainfall (ER1(t) and ER2(t)) is produced depending on the current catchment moisture state described by the storage capacity distribution function F(c). The parameter CMAX describes the maximum storage capacity. The effective rainfall is distributed by the parameter ALPHA and either routed through three linear reservoirs with residence time Kq in series, or a single reservoir with residence time Ks. (From Wagener *et al.*, 2001; reproduced with permission of the European Geosciences Union)

voir for the slow response) in parallel as a routing component (Fig. 4.10). The SMA component is described in detail in Section 3.2.3 (though the notation of Wagener *et al.*, 2001 is used here) and by Moore (1985, 1999). The component assumes that soil moisture capacity varies across the catchment and therefore that the proportion of the catchment with saturated soils varies over time. The spatial variability of soil moisture capacity is described by the following distribution function:

$$F(c) = 1 - \left(1 - {}^{c(t)}/_{CMAX}\right)^{BEXP}, \quad 0 \le c(t) \le CMAX \tag{4.2}$$

The structure requires the optimization of five parameters: the maximum storage capacity in the catchment CMAX [L]; the degree of spatial variability of the soil moisture capacity within the catchment BEXP [-]; the factor distributing the flow between the two series of reservoirs ALPHA [-]; and the coefficients of the linear reservoirs Kq [T^{-1}] and Ks [T^{-1}]. Actual evapotranspiration is equal to the potential value if sufficient soil moisture is available; otherwise it is equal to the available soil moisture content.

4.4.1.3 Methodology

Traditional automatic calibration schemes use single value objective functions such as the RMSE (see Eq.on 4.1). In this example, however, a partitioning scheme suggested by Boyle *et al.* (2000) to define OFs based on the different response modes of the hydrological system is utilized. The approach is based on the reasonable assumption that the behaviour of the catchment is inherently different during periods *driven* by rainfall and periods without rain. Further, the periods immediately following the cessation of rainfall and dominated by interflow can be distinguished from the later periods that are dominated by baseflow. The streamflow hydrograph can, therefore, be partitioned into three components (Fig. 4.11): *driven* (Q_D), *non-driven quick* (Q_Q), and *non-driven slow* (Q_S).

Time steps corresponding to each of these components are identified through an analysis of the precipitation data and the time of concentration for the catchment. Time steps with non-zero rainfalls, lagged by the time of concentration for the catchment, are classified as driven. Of the remaining (non-driven) time steps, those with streamflow lower than a certain threshold value (e.g., the mean of the logarithms of the flows) are classified as non-driven slow, and the rest are classified as non-driven quick. The model performance during these three periods (Q_D, Q_Q, Q_S) is estimated by calculating the RMSE (FD, FQ, FS) separately over each period.

The primary motivation for partitioning the non-driven flows into a quick and a slow component is to identify the periods of hydrograph recession or baseflow behaviour from the rest of the non-driven flow. For the purposes of this study, a simple systematic approach (threshold flow value) was chosen to identify these periods. The sensitivity of the threshold values to the identification of the recession periods was investigated prior to the multi-criteria optimization. Several different threshold values were tested (median of flows, mean of flows, mean of log of flows, etc.) to determine which value provided the best representation of the recession flows as determined through visual inspection of the observed hydrograph (results not shown here). The mean of the log of the flows provided the *best* estimate of the recession periods for this data set. There are certainly other possibly more accurate methods (e.g., visual inspec-

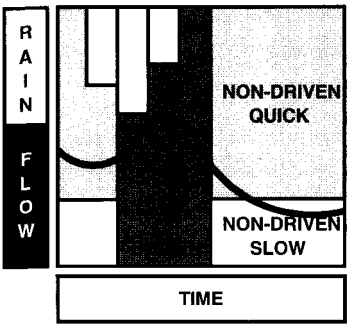


Fig. 4.11 Hydrograph segmentation into three components based on different response modes of the catchment system, i.e., 'driven' (QD - dark grey), 'non-driven quick' (QQ - light grey) and 'non-driven slow' (Q_S - white) flow. (From Wagener *et al.*, 2001; reproduced with permission of the European Geophysical Union)

tion, water balance, and groundwater recharge methods) to identify these recession periods; however, these have to be the subject of future studies. Presumably, the more accurately the characteristic features of the catchment are identified, the more informative the analysis.

Two calibration methods, uniform random search (URS) and the multi-objective complex evolution algorithm (MOCOM, Yapo *et al.*, 1998), were used to explore the parameter space of the model. The URS method consisted of 5000 parameter sets randomly sampled from the feasible parameter ranges, based on a uniform distribution. The Pareto optimal solution space for the three criteria was estimated with 500 solutions using the MOCOM multi-criteria optimization algorithm. In brief, the MOCOM method involves the initial selection of a population of p points distributed randomly throughout the s-dimensional feasible parameter space. In the absence of prior information about the location of the (Pareto) optimum, a uniform sampling distribution was used. For each point the multi-objective vector F was computed, and the population was ranked and sorted using a Pareto-ranking procedure suggested by Goldberg (1989), *i.e.,* within the population of a certain rank it was not possible to find a parameter set that is better than another with respect to all objective functions. Simplexes of $s + 1$ points were then selected from the population according to a robust rank-based selection method (Whitley, 1989). A multi-objective extension of the downhill

simplex method (Nelder and Mead, 1965) was used to evolve each simplex in a multi-objective improvement direction. Iterative application of the ranking and evolution procedures caused the entire population to converge towards the Pareto optimum. The procedure terminated automatically when all points in the population become non-dominated, *i.e.*, of rank one. Experiments conducted using standard synthetic multi-objective test problems have shown that the final population provides a fairly uniform approximation of the Pareto solution space (Yapo *et al.*, 1998; Bastidas, 1998).

4.4.1.4 Results and discussion

A modification of the RSA approach introduced by Freer *et al.* (1996) was used to visually inspect the sensitivity of the different parameters with respect to the response mode of the system (see Section 3.3.3). This methodology was introduced originally to identify insensitive parameters that subsequently would be fixed or eliminated. However, it can also be used to visualize the link between parameter sensitivity and system response modes (Dunne, 1999; Wagener *et al.*, 1999). Freer *et al.* (1996) split the parameter population, derived from a URS procedure and ranked with respect to their OF values, into ten groups of equal size and plotted the cumulative distribution of the parameters in each group with respect to the transformed measure of performance. The measures were transformed so that higher values indicated better models and were divided by their sum so as to add up to unity. An insensitive parameter would produce a straight line, while differences in form and separation of the resulting curves indicate parameter sensitivity. Splitting the population into ten groups, instead of just two as in the original method, avoided the selection of a threshold value between behavioural and non-behavioural parameter sets, and increased the information gained by the analysis. Figure 4.12 visualizes the results derived for this study with the shading ranging from light grey (best performing group) to black (worst performing group). The figure shows the sensitivity of the model parameters based on the RMSE, first as an overall measure for the whole calibration period (first row), and subsequently as a measure for the three different response modes (FD, FQ, FS). The overall RMSE and the FD

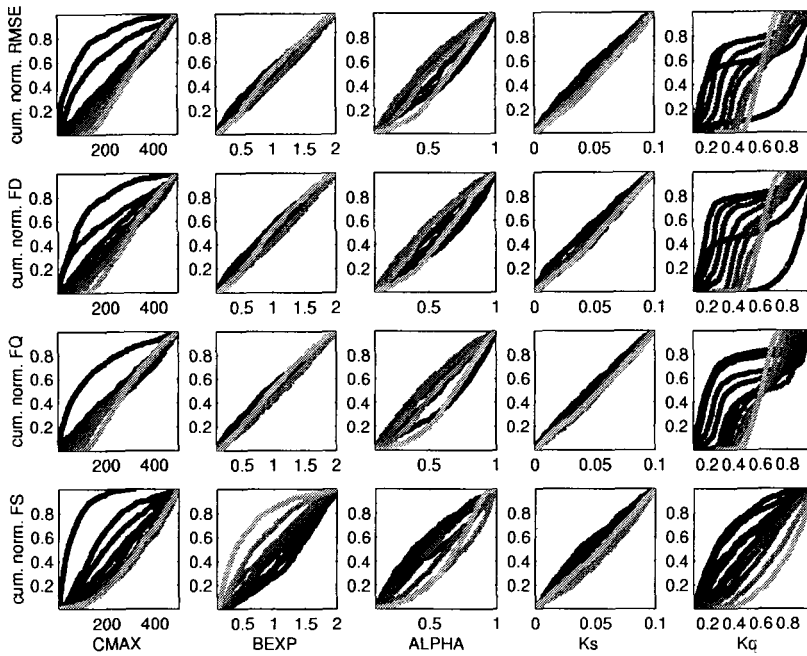


Figure 4.12 Regional Sensitivity Analysis plots showing the varying sensitivity of the model parameters when using different objective functions. (From Wagener *et al.*, 2001; reproduced with permission of the European Geosciences Union)

measures show very similar behaviour, indicating that they retrieve similar information from the observed data.

The curves produced using these two measures are markedly different from those resulting from the FS measure. The sensitivity of the BEXP parameter is considerably higher during periods of non-driven slow response, *i.e.*, FS. The sensitivity of Kq is relatively high for all measures. However, the shape of the cumulative curves of this parameter for the FS measure is different. This indicates that the parameter population conditioned on this measure results in a different distribution than when the other measures are used. Sensitivity plots for the parameter ALPHA are similar for all OFs, suggesting that this parameter is equally important for the correct reproduction of the system behaviour during all response modes. The same is observed for parameter CMAX.

Parameter sensitivity is a necessary but not sufficient requirement for identifiability, since values of a sensitive parameter that produce good model performance can still be distributed over a relatively wide range of the feasible parameter space. A simple way to show how parameter identifiability is increased through the use of different measures is demonstrated in Fig. 4.13. The parameter populations (derived from the URS) were transformed as before so that again the OF value of the best performing parameter was assigned the highest value and all measures sum to unity. The range of each parameter was then split into M containers and the sum of all measures in each was calculated. The results are the densities of the (in this case, initially uniform) parameter populations conditioned on the different objective functions.

Some variation in the distributions derived through the use of the different measures can be seen. The parameter BEXP shows relatively uniform distributions except when being conditioned on FS, where small values show better performance. The parameter population of Kq on the other hand shows a very distinct peak for the FQ objective function. However, higher values of this parameter are favoured when the FS objective function is used in order to fit the beginning of the recession periods of the hydrograph better. ALPHA shows relatively similar distributions for all measures, a result that is in line with the sensitivity analysis (Fig. 4.12) in which the parameter is sensitive for all objective functions.

Figure 4.13 shows that the use of different measures can lead to an improvement in judging the performance of a parameter over its range. A parameter showing little variation using one measure may reveal a distinct peak in its distribution when using an objective function based on the residuals from a different response period. This is caused by the varying importance of different model components (and therefore different parameters) to reproduce the system behaviour during different response modes.

A two-dimensional projection of the three-dimensional objective function space (FD, FQ, FS) gives further insights (Fig. 4.14, first two rows). Light grey dots indicate the 500 Pareto solutions determined with the MOCOM algorithm whereas the black dots show the 5000 URS results. The second row shows the region of the Pareto solution in greater detail with the best solutions highlighted (A for FD, B for FQ, C for FS,

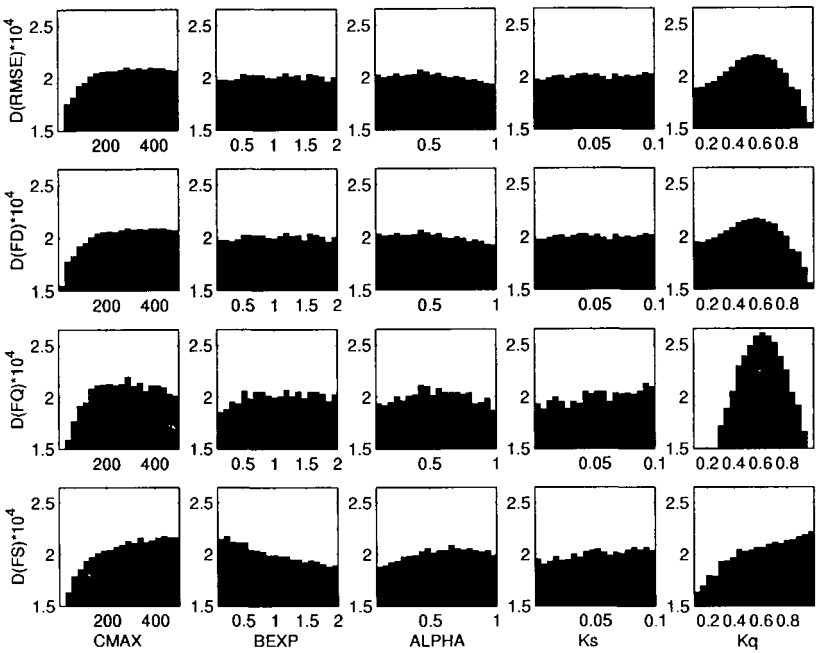


Fig. 4.13 The objective functions are rescaled so that the best performing parameter assumes the highest value and the sum of all values equals one. Splitting each parameter range subsequently into 20 bins of equal width and calculating the sum of all measures in each container leads to the parameter density (D) distributions shown. (From Wagener *et al.*, 2001; reproduced with permission of the European Geosciences Union)

D for overall RMSE). These plots clearly illustrate the inability of the model to simultaneously match all three aspects of the hydrograph, and reveal that the trade-offs in fitting the three-hydrograph components are quite significant. However, the trade-off between FD (A) and FQ (B) is relatively small, as is also indicated by the relatively high degree of correlation of the Monte Carlo results (top left plot). In addition, the best FD and overall RMSE solutions are very similar with respect to the three criteria, again indicating that the two measures contain very similar information about the parameters of this model. The normalized (over the initial parameter uncertainty range) parameter plot, presented in the bottom row of Fig. 4.14, shows the variability in the parameter values for the 500 Pareto optimal solutions (indicated by the light grey lines). Each line on the graph represents one of the parameter sets. Notice that the pa-

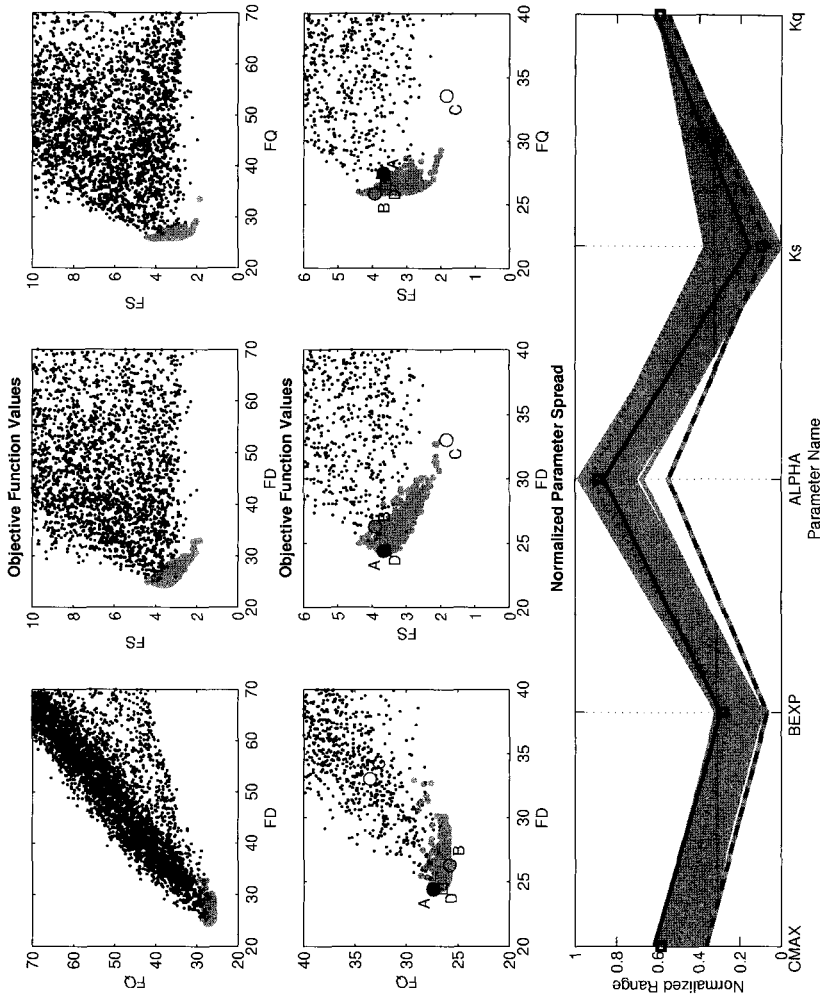


Fig. 4.14 Two-dimensional projections of the three dimensional objective function space (1st and 2nd row show 500 Pareto solutions and 5000 parameter sets randomly sampled from a uniform distribution). The markers correspond to the best points with respect to FD (A), FQ (B), FS (C), and overall RMSE (D). The 3rd row shows the normalized parameter space. The grey lines show the 500 Pareto solutions, the three black lines are solutions A (FD, solid), B (FQ, dotted), and C (FS, dashed). The squares indicate the best overall RMSE solution (D) (Figure by Douglas P. Boyle, from Wagener *et al.*, 2001; reproduced with permission of the European Geosciences Union)

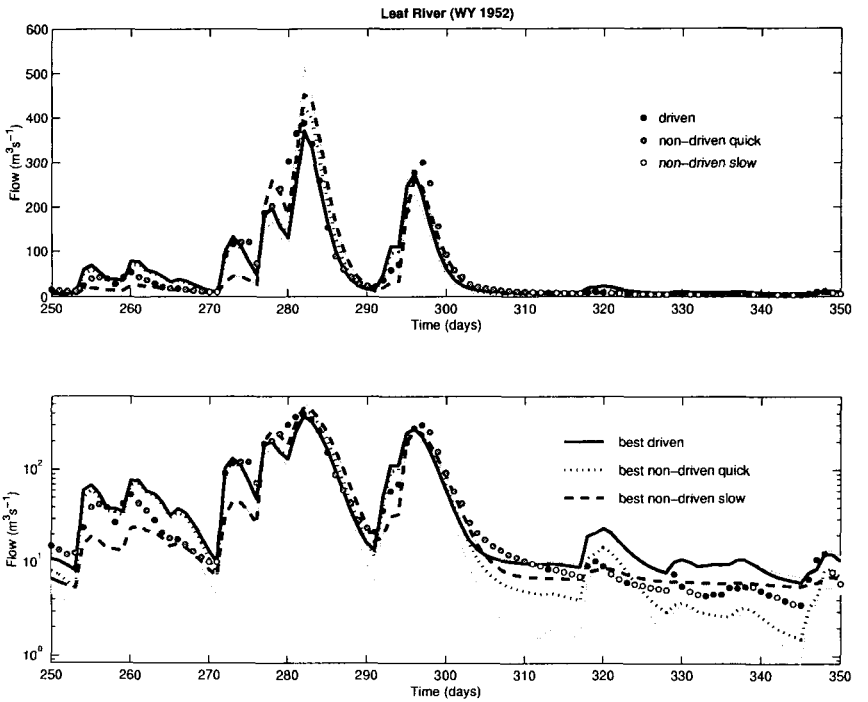


Fig. 4.15 Hydrograph range produced by the 500 Pareto solutions (grey region), and the output from the best parameter sets for the different measures on normal (1st row) and on logarithmic scale (2nd row). The observed time-series is shown as circles. (Figure by Douglas P. Boyle, from Wagener *et al.*, 2001; reproduced with permission of the European Geosciences Union)

rameter uncertainty has been reduced significantly by the multi-criteria optimization compared to the initial feasible range, particularly for Kq. Also notice that the parameter values for the best FD, FQ, and RMSE solutions are, in general, in a different region of the parameter space than the best solution for the FS criteria (indicated by the dashed line).

Figure 4.15 presents the model output results for a 100-day portion of the calibration period derived using the results of the calibration with the MOCOM algorithm. The minimal FD and FQ solutions tend to fit the peaks better at the expense of over- and underestimating the recessions respectively. The minimal FQ solution also captures the shape of the falling limb, corresponding with time steps classified as *non-driven quick*

better than the other two solutions. The minimal FS solution on the other hand fits the long recession limbs of the hydrograph better (see log-scale plot at bottom), while it often seriously over- or underestimates the peaks. The model generally has some trouble matching the flows for days 250 through 270. This could be due to model structural error, the model's inability to track the soil moisture in the long dry period preceding these rainfall events. Another possibility is that the precipitation data during this time period is erroneous, *i.e.,* it may not be representative of the precipitation rates throughout the catchment.

This simple example demonstrates how the aggregation of the residuals over the whole calibration period results in a loss of information relating to parameter sensitivity and identifiability, model performance, and model structural insufficiencies. Additional insight is gained from the hydrograph split performed here.

The advantages of a multi-objective framework based on system response modes make it especially suitable for comparison studies since it allows the modeller to attribute the model performance during different system response modes to different model components, in this case the moisture accounting and the routing components. A certain model structure might perform better during *driven* periods because of a superior moisture accounting component, while another model structure containing a more appropriate slow flow routing component could result in a higher performance during *non-driven slow* periods. A single-objective framework does not allow the comparison of model components and consequently important information relevant to identifying the most suitable model structure is lost.

A second model structure is introduced here to further demonstrate the advantage of multi-objective comparisons. This structure is a simplified version of the model described before. A single store with a rainfall excess mechanism is used as the moisture accounting component, instead of a distribution of stores as in the first structure. The store is described by its size, CMAX [L]. The evapotranspiration losses of the store are again equal to the potential rate as long as soil moisture is available. The remaining components are identical to the ones in the earlier model structure shown in Fig. 4.10 and are defined by three parameters, AL-

PHA [-], Kq [T^{-1}], and Ks [T^{-1}]. This structure is referred to as the simple model, while the initial structure is referred to as the complex model in the remainder of this section.

The results of the comparison are shown in Fig. 4.16. The left graph shows a comparison in performance between the two structures as derived from the calibration with the MOCOM algorithm. The OFs used are identical to the ones applied when analysing the individual model structure earlier in the text. The traditional, overall measure of performance, RMSE, indicates that the complex model structure is superior to the simple one. However, when analysing the performance in more detail one can see that both structures reproduce the non-driven slow periods (FS) almost equally well. The complex model structure is better able to fit the driven and non-driven quick periods, with the largest difference occurring during the driven period. This result shows the additional insight gained by a more detailed analysis. The model structures have identical components to fit the slow catchment drainage (Q_S) and therefore in this case produce similar results. The fact that this is not the case during

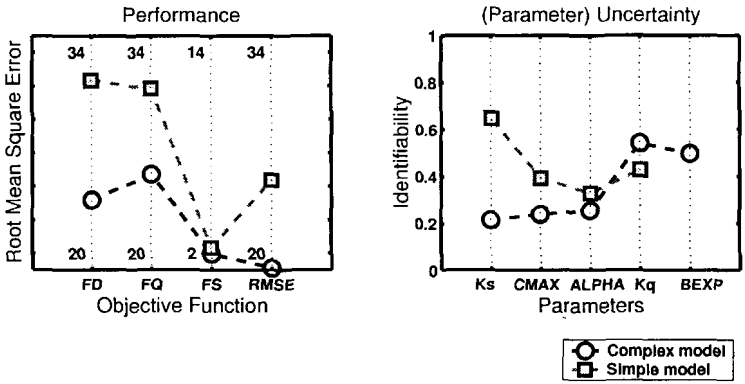


Fig. 4.16 The two model structures compared in terms of performance and uncertainty in (identifiability of) their parameters. The left plot shows the root mean square error values of the two structures with respect to the different objective functions used. A smaller value therefore indicates a higher performance. The plot on the right shows the identifiability of the parameters of the two structures. A higher value indicates a higher degree of identifiability and therefore reduced uncertainty. The identifiability value for each parameter is the highest derived from the different objective functions. (From Wagener *et al.*, 2001; reproduced with permission of the European Geosciences Union)

the quick catchment drainage (Q_Q) can be attributed to the larger importance of the moisture accounting component in fitting this part of the hydrograph, and the fact that it is easier to separate out the slow recession periods.

The right-hand graph of Fig. 4.16 shows the uncertainty in estimating the parameters of the two models in terms of their identifiability, using the measure introduced in Fig. 4.3. The highest identifiability values for each of the parameters of the complex model are used. Identifiability values for the simple model parameters are derived in an identical way. It can be seen that the simple model structure shows an overall higher degree of parameter identifiability. Introducing an additional parameter, BEXP, reduced identifiability through its interaction with the other parameters. The increase in model performance is therefore obtained at the cost of decreasing identifiability, and increasing parameter uncertainty.

The trade-off between improvement in performance and reduction in identifiability should be considered, among other things, when selecting a model structure for a specific purpose. The type of analysis shown supports this type of model selection process.

4.4.2 Example 2 – Dynamic identifiability analysis

4.4.2.1 Data

The river selected for this study is the Lower Medway at Teston (1256.1 km^2) located in South Eastern England. Six years of data (10/04/90 – 14/07/96) of daily naturalized flows, precipitation, PE and temperature are available (Fig. 4.17). The Medway catchment is characterized by a mixture of permeable (chalk) and impermeable (clay) geologies subject to a temperate climate with an average annual rainfall of 772 mm and an average annual PE of 663 mm (1990–1996). Two measures of the Hydrology of Soil Types classification (HOST, Boorman *et al.*, 1995; see Chapter 5) for the UK indicate the response character of the catchment. These are the Base Flow Index (BFIHOST) and the Standard Percentage Runoff (SPRHOST), which are 0.439 and 41.4 respectively.

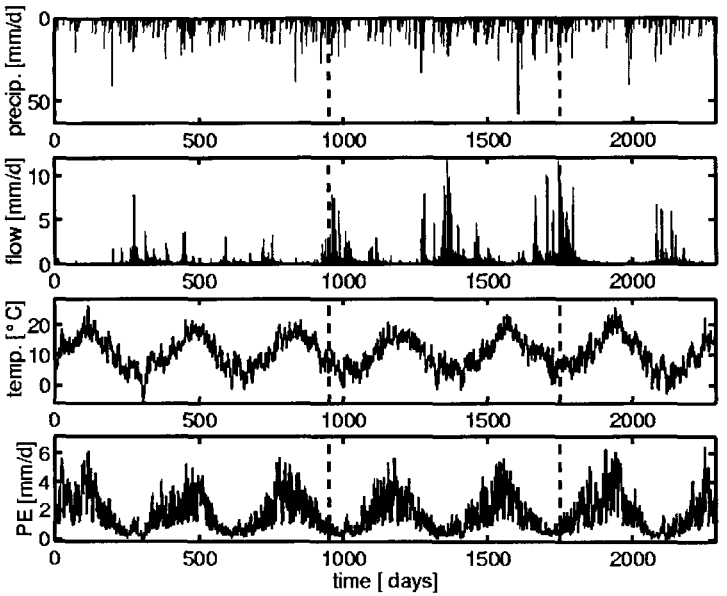


Fig. 4.17 Data for the Medway at Teston for a six-year period (10/04/1990 – 14/07/1996). Dashed lines enclose period used for dynamic analysis.

4.4.2.2 Model structure

The structure selected for analysis is a combination of a Penman type SMA component (Penman, 1949), as used by Jolley (1995), and a parallel routing structure consisting of two linear conceptual reservoirs to represent quick and slow catchment response (Fig. 4.18). The structural elements were introduced in Chapter 3, but the details are repeated here for convenience. The ratio of flow contributing to each of the two routing reservoirs is fixed. The model structure contains five parameters in total. The Penman model has two parameters: the size of the near surface store, defined by a root constant *rc* plus an additional 25mm to allow for capillary rise (Penman, 1949), and a bypass parameter. The bypass component represents phenomena that divert water from the soil moisture store and lead to rapid groundwater recharge or runoff response during rainfall, such as macropore and infiltration excess overland flow (Jolley,

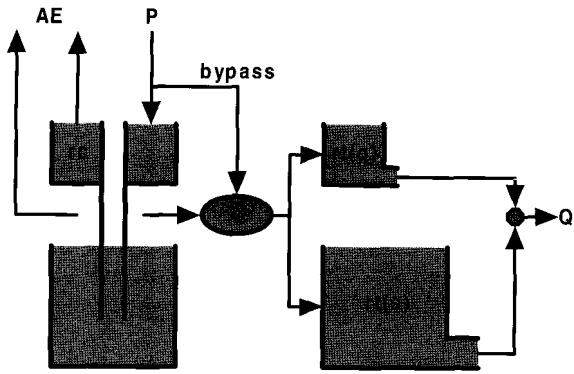


Fig. 4.18 Model structure applied. (From Wagener *et al.*, 2003a; reproduced with permission of John Wiley & Sons, Inc.)

1995). It applies to the proportion of rainfall that exceeds the potential evapotranspiration. The near surface store is connected by an overflow mechanism to the lower store. The size of the lower store is chosen large enough to ensure that it never empties (Moore, 1999). Additional effective rainfall is produced when both stores are filled and the lower store overflows. Evapotranspiration takes place at the potential rate from the near surface store. It reduces to 1/12 of the potential rate from the lower store when the upper store is emptied, as suggested by Penman (1949). The split of the effective rainfall between quick and slow flow is defined by the parameter alpha, which is the ratio of flow going toward the quick response reservoir. The remaining two parameters are the residence times of the two linear stores rt(q) and rt(s).

This structure was selected because is contains components that can be found in many CRR model structures, e.g., a two layer soil moisture accounting component producing effective rainfall (e.g., Greenfield, 1984; Jolley, 1995), and a routing component consisting of two parallel stores with a fixed flow distribution between them (e.g., Jakeman and Hornberger, 1993; Sefton and Howarth, 1998; Young and Beven, 1994).

4.4.2.3 Results and discussion

Traditional Monte Carlo sampling

The result of a conventional Monte Carlo uniform sampling of 20,000 points in the feasible parameter space is shown in the form of dotty plots in Fig. 4.19. This is used as a benchmark for evaluating the DYNIA results. The OF used in Fig. 4.19 is the RMSE, in this case using the residuals over the whole six-year period. It can be seen from these plots that some of the parameters show quite a distinct optimum (e.g., rt(q) or alpha), while others (in particular rt(s)) reveal equally performing values over a relatively widespread part of the feasible parameter space.

Information content

The first step in the DYNIA is to separate periods of high and low information content with respect to each of the parameters. The information-rich periods can then be used in various ways, for example linked to specific response modes of the natural system or used to define parameter (group) specific OFs.

The information content is calculated using the first two steps of DYNIA, shown in Fig. 4.6(a) and (b). The cumulative distributions calculated for every time step (Fig. 4.6[b]) can be used to derive confidence limits for the different parameters. Wide confidence limits suggest that parameter values associated with equally good performance are distributed widely over the parameter space; narrow limits suggest that the best performing parameters are concentrated in a small area of the feasible range. A transformed measure (one minus the width of the confidence limits over the parameter range, normalized to run from 0 to 1) is used here so that a large value is equal to a high information content for a given time-step. The time-series of the information measure is plotted for each of the parameters in Fig. 4.20, together with the streamflow (normalized for display) and the rainfall. It is important to remember that this plot contains a subjective element through the definition of the initial feasible parameter space by the modeller.

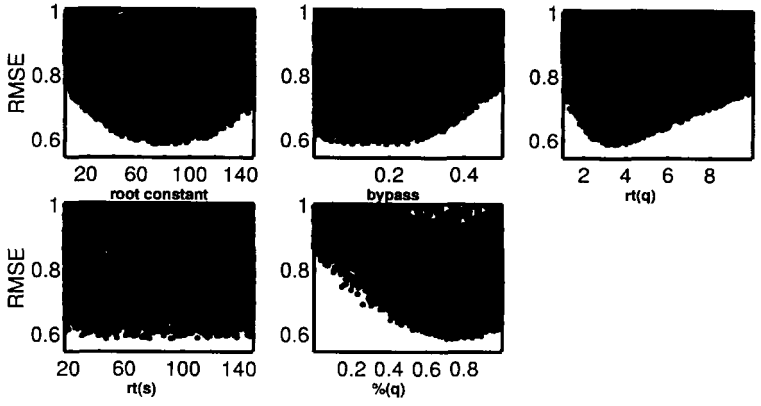


Fig. 4.19 Dotty plot showing results of the uniform random search using 20000 samples utilizing the whole six-year period available. Lower values of the Root Mean Squared Error (RMSE) objective function indicate better performing parameter values. Only parameter sets producing an RMSE below 1 mmd^{-1} are shown. (From Wagener *et al.*, 2003a; reproduced with permission of John Wiley & Sons, Inc.)

Figure 4.20(b) shows the information content for the root constant, *rc*, derived using a window size of 101 (daily) time steps. It can clearly be seen that the main information about the root constant emerges toward the end of long recession periods (dry summers) and in particular during the wetting up periods. The information values during the remaining time are relatively small.

The bypass parameter is analysed in Fig. 4.20(c) with a window of 41 time steps. The plot reveals three types of periods where information is available. The first is for small runoff events after wet winter periods, e.g., around days 150 and 500; the second is located in summer periods, e.g., days 175–300 and 550–700; and the third during wetting up phases, e.g., days 350 and 725.

Information about the quick flow residence time, rt(q) (Fig. 4.20[d], using a window of 11 time steps), can mainly be found during the quick recessions after high flow events, while the long recession tails contain the information about the slow flow residence time, rt(s) (Fig. 4.20[e],

using a window of 41 time steps). Using larger window sizes for rt(s) did not improve the result. An attempt at using a regressive variant of the moving window approach to improve the results for the residence time parameters, in which only a certain number of time steps up to the time step itself are considered, was not successful. This is because, especially when large window sizes are used, periods of high identifiability are shown after the time steps which actually contain the information.

The analysis of the split parameter alpha (Fig. 4.20[f], using a window size of 21 time steps) reveals that this parameter becomes identifiable after flow events when the response is changing from quick to slow drainage. Little information about this parameter can be gathered during periods of long recessions that have only minor runoff events.

The information contained in Fig. 4.20 can also be used to find combinations of parameters responsible for the model's behaviour during specific response periods. These interacting parameters could then be grouped for multivariate calibration (e.g., Wheater *et al.*, 1986). A threshold for the information content value of 0.3 is, somewhat arbitrarily, selected here, and the selected high information content time steps for the different parameters are shown in Fig. 4.21 (each parameter is indicated by a different grey shading) together with the normalized streamflow. From this plot it is easy to see that rt(q) and alpha show high information contents during similar periods, and therefore during the same response modes. There is also a considerable overlap between the relevant periods for the bypass and rt(s) parameters, at least during the first slow recession phase. However, these similarities do not necessarily imply parameter interdependence.

The initial Monte Carlo simulation over the whole calibration period was based on the RMSE. It is clear from Fig. 4.19 that this measure is not capable of retrieving information to distinguish between the performances of different values of rt(s), which only become identifiable during distinctive periods of recession. The RMSE emphasizes performance during peak flow periods. However, applying a simple threshold to the data, separating out periods of low flow, can improve identifiability as demonstrated in Fig. 4.22. The three different lines display the gradient distributions over the range of rt(s). The gradients are derived using steps (a) and (b) of the DYNIA procedure (Fig. 4.6), where the feasible range

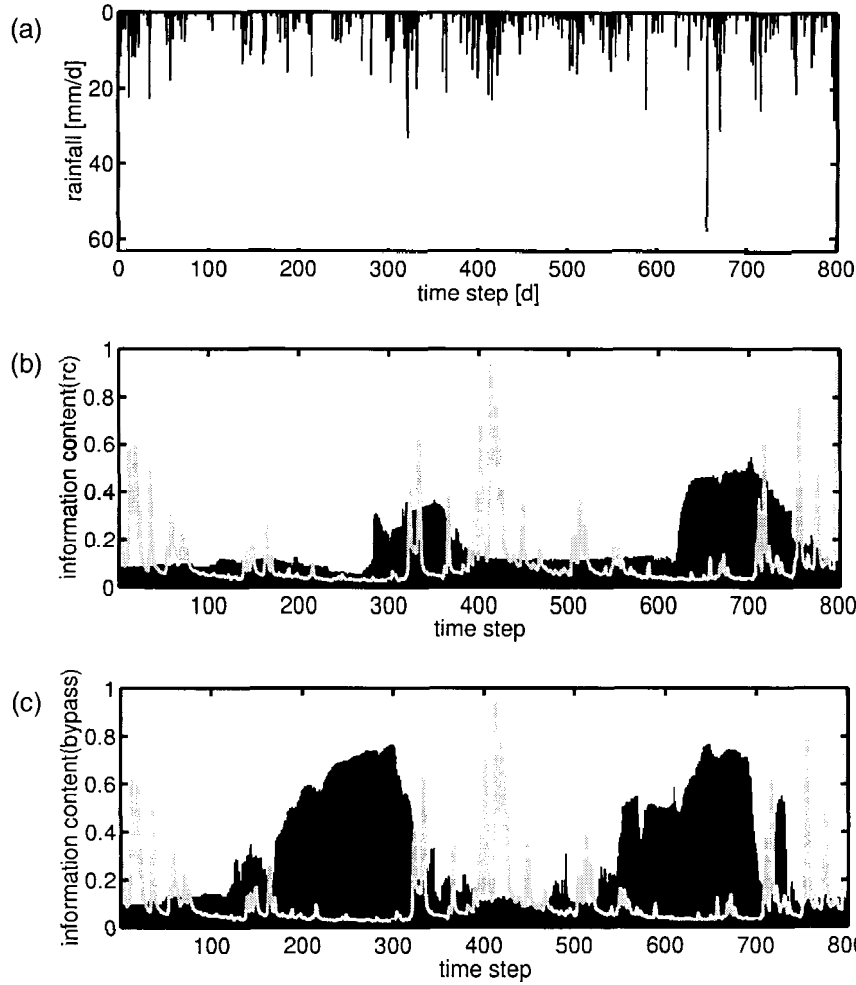


Fig. 4.20 The information content of the data over a two-year period (days 950 to 1750). Graph (a) shows the precipitation input over the considered period. The remaining plots show the result for the different parameters (black bars). The grey line is the streamflow, normalized with respect to its maximum value. (From Wagener *et al.*, 2003a; reproduced with permission of John Wiley & Sons, Inc.)

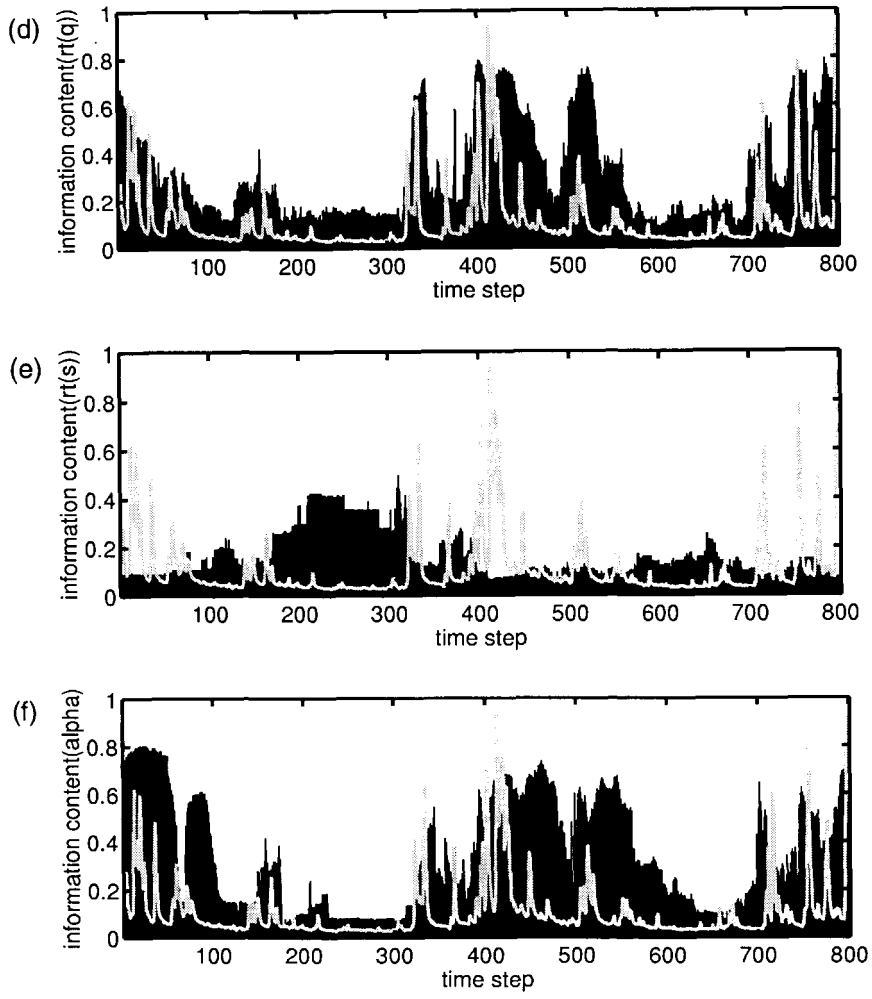


Fig. 4.20 (cont.)

is split into ten containers of equal size. These gradient distributions represent the full data record (full line), and those time steps where the observed flow is below a certain threshold (dotted line), *i.e.*, 0.5 mmd^{-1}, in order to consider only periods of recession. It can be seen that the identifiability is improved. The dashed line shows the result when only time steps below the selected threshold and with an information content of

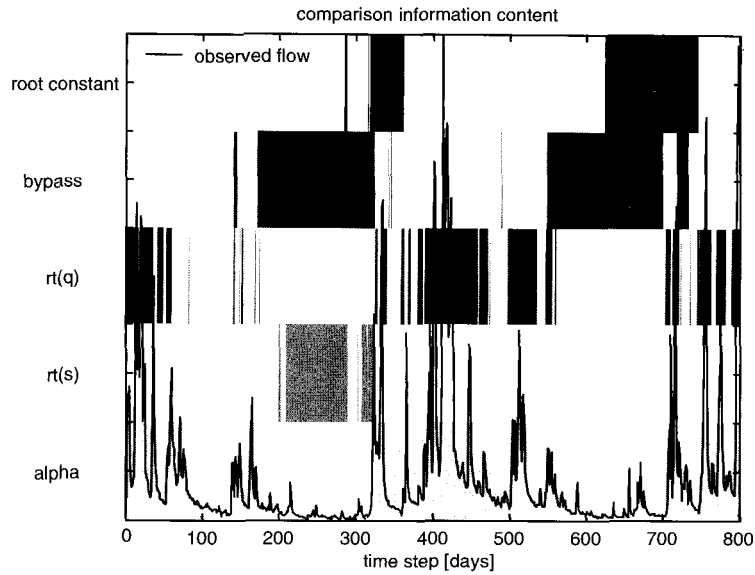


Fig. 4.21 Comparison of the information content of a two-year period for the different parameter values (days 950 to 1750). Only time steps with an information content above 0.3, with respect to individual parameters, are shown in the different grey shades. (From Wagener *et al.*, 2003a; reproduced with permission of John Wiley & Sons, Inc.)

above 0.15 are considered. The additional flow criterion is required since informative regions can also be found during high flows that show different optima (*i.e.*, very low values) in the parameter space (full line). This is caused by structural inadequacies in the simple slow flow component of the model. The dashed line shows the highest identifiability values and reveals that the best performing values for rt(s) increase in magnitude when the influence of high flow periods is removed.

Dynamic identifiability and structural failure

Information content plots only describe where in the time series a parameter becomes identifiable. They do not give any information about the location of optima in the parameter space. A different type of plot is therefore shown in Fig. 4.23, derived by performing the remaining stages of the DYNIA. The plots visualize the DYNIA results in the parameter-time space. The values of the identifiability measure values are trans-

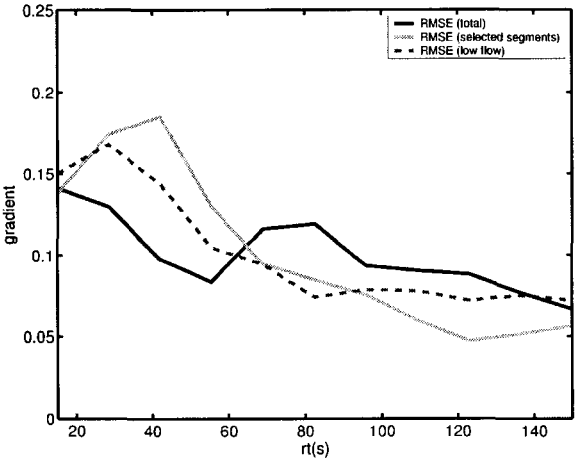


Fig. 4.22 Comparison of identifiability, defined as the gradients of the cumulative distribution at different locations of the parameter range, of the slow response residence time rt(s) using the RMSE (1) as an overall measure of the performance over the whole calibration period (continuous line); (2) only utilizing the residuals at time steps with flow values below 0.5mmd^{-1} (dotted line); and (3) using residuals with flow values below 0.5mmd^{-1} and with an information content above 0.15 (dashed line). (From Wagener *et al.*, 2003a; reproduced with permission of John Wiley & Sons, Inc.)

formed into grey shading, with higher values indicated by a darker colour, and plotted against the time axis (see Fig. 4.6[c] and [d]). Additionally the 90% confidence limits (derived from the cumulative distributions) and the streamflow (normalized with respect to its maximum value) are shown.

Figure 4.23(b) shows the results for the root constant, *rc*. It can be seen that the confidence limits narrow during the wetting up periods after dry summers. During those periods the parameter clearly strives to higher values. No particular optima are visible during the remaining periods, indicating that very different values of this parameter yield similar results in combination with the remaining parameters. There are however two different optima visible in Fig. 4.23(b), one value of approximately 80 and the other around 120. The optima jumps between those two values. This indicates that the optimum values will be influenced by the particular local conditions (in terms of time). For parameter estimation one

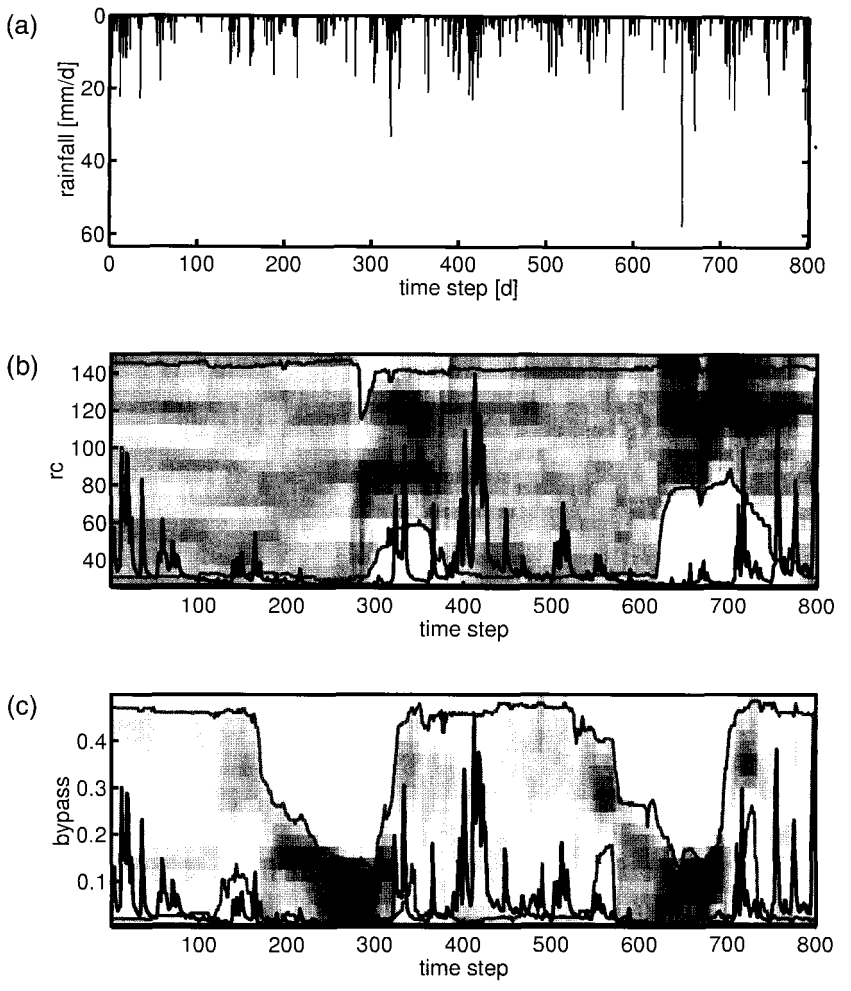


Fig. 4.23 Results of the DYNIA procedure for a two-year period (days 950 to 1750). Graph (a) shows the rainfall input over this time. The remaining graphs show the DYNIA results for the different parameters. The grey shading indicates the size of the gradient, with a darker colour for a higher value. The dark grey lines are the 90% confidence limits derived from the cumulative distribution of support values, while the black line is the streamflow normalized with respect to its maximum value. (From Wagener *et al.*, 2003a; reproduced with permission of John Wiley & Sons, Inc.)

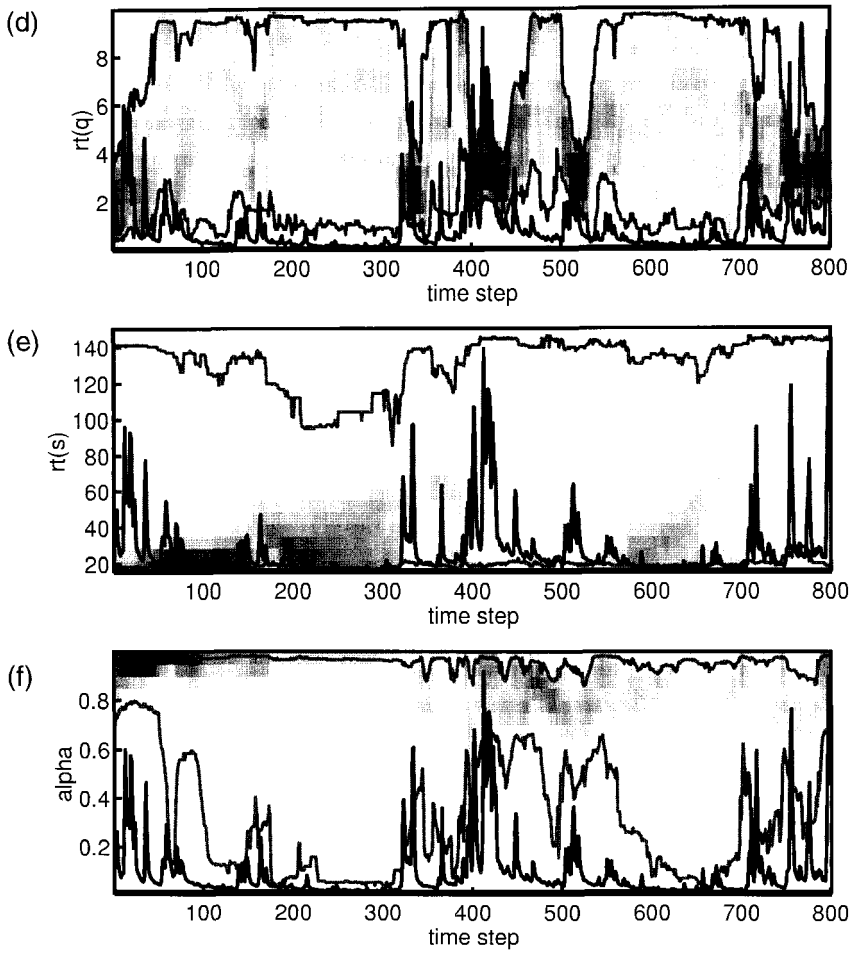


Fig. 4.23 (cont.)

would have to take an average value over different periods to account for this effect. This parameter requires the largest window for its analysis. Investigations so far suggest that this is typical for parameters that describe the maximum storage capacity.

The bypass parameter has different periods showing a high information content as shown in Fig. 4.20(c). In Fig. 4.23(c), one can see that the parameter jumps between at least two optima. Small values of this parameter perform better during low flow periods (e.g., around time steps

250–300 and 650–700). During other periods, e.g., around time-steps 150, 550 and 725, larger values of the bypass (0.3 or higher) seem to provide a better fit to the observed data, albeit with decreased identifiability (lighter grey shading). Small bypass values are required during summer periods to yield only little runoff during summer storms when the catchment is dry. High values are needed during storm events after wet periods and in the wetting up periods. This can explain why the Monte Carlo simulation results shown in Fig. 4.19, using the RMSE, provide no clear optimum with respect to this parameter. The RMSE error measure is biased toward fitting higher flow values and the high identifiability areas during slow recessions might not be influential enough to produce a clear peak on the dotty plots. This change in optimum parameter value within the parameter space is an indicator of failure of the model structure. There is a clear inconsistency in the way the model fits the observed behaviour of the catchment.

Parameter rt(q) is analysed in Fig. 4.23(d). Clear optima occur at a value of approximately 3 during the quick recession periods, while there are no specific peaks during other time steps.

It is at best very difficult to identify suitable values for rt(s) using a measure like RMSE. As noted above, the residuals of slow flow periods are often too small to influence the overall performance of a model, as shown in Fig. 4.19. Figure 4.23(e) however, shows that better performing values for this parameter lie near to the lower boundary of the feasible parameter space, especially at time-steps 200–300, which is the longest recession period contained in the data-set used.

Parameter alpha does not show such a distinct area of identifiability as for example *rc* or bypass, apart from a short beginning period within the warm-up range of the algorithm (Fig. 4.20[f]). There is however some area of darker grey shading in the period between time steps 400 to 550, which is after the main wet period. The parameter varies roughly between 0.6 and 0.9 during this period, the range also found in the initial Monte Carlo analysis (Fig. 4.19). Figure 4.20(f) suggests some variation of this parameter in time. However, the evidence here is not sufficient to draw conclusions and further analysis is required.

Two-dimensional response surface plots are used in Fig. 4.24 to analyse the identifiability and interaction between the soil moisture account-

ing parameters, *i.e.,* the root constant and the bypass, more closely. The upper plot shows the result when the RMSE is calculated using the residuals for the complete time-series, while the bottom plot only uses time-steps with high information content and excludes those that show ambiguity with respect to the bypass parameter. One can see that the parameters are much better identified when periods of high noise are not considered. The optimum values for the root constant are however slightly smaller than suggested in the DYNIA plots. This emphasizes the need for a detailed analysis using different methods of visualization with different degrees of detail.

Inference and areas of possible model structure improvement

The analysis performed using DYNIA indicates different failures of the model structure with respect to the underlying assumptions. These failures can be used to suggest areas of improvement. However, while the identification of a failure is relatively straightforward and objective, the resulting course of action is not. The analysis of why a failure occurred and how the model structure can be improved very much depends on the experience and creativity of the modeller himself. As Beck (1985) points out, there is no systematic *algorithm* for changing an inadequate structure that is equivalent to increasing a polynomial order from n, say, to $(n + 1)$, as would be possible for some data-based model structures. The modeller's task is to draw inference from the type of failure that has occurred with respect to the hypothesis underlying the specific model component in order to develop an improved version.

For example, the structural failure implied by the two distinct regions of preferred values of rc (Fig. 4.23[b]) could be related to the inflexibility of this component in fitting the (dynamic) wetting up period at the end of dry summers. A more adaptable variant of this component might perform better, for example using a probability distribution of moisture stores in the catchment (Moore and Clarke, 1981; Moore, 1999).

The analysis also indicates a failure with respect to the bypass parameter. This parameter shows distinct areas of well-performing values in different parts of the feasible parameter space, as described earlier.

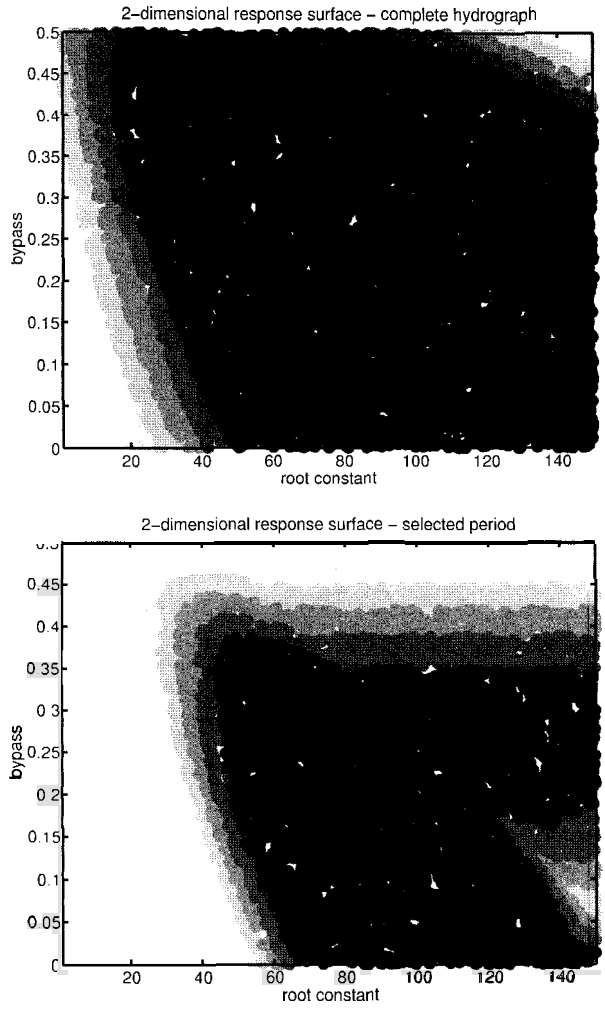


Fig. 4.24 These two plots represent the response surface between the two parameters of the soil moisture accounting component, the root constant and the bypass. Both are based on a uniform random search sampling 10,000 points, during which the routing parameters were fixed to well performing values. Both plots consist of individual dots. The white areas are caused by a lack of density. The RMSE in the top plot is calculated using the residuals from the complete time-series, while the bottom plot uses periods of high information content (see Fig. 4.21), while avoiding the ambiguity of the bypass parameter identified in Figure 4.23. The time-steps used in the selected period are 200 to 375 and 600 to 750. The smallest RMSE values are shown in black and the values increase by steps of 0.05mmd^{-1} per contour. (From Wagener *et al.*, 2003a; reproduced with permission of John Wiley & Sons, Inc.)

Reasons for this could be that the process represented by the parameter is more complex than assumed, or that different processes, which could be represented separately, are aggregated into a single component. A possible improvement would be the replacement of this constant parameter with a dynamic component, as by making the amount of precipitation that contributes directly to the effective rainfall dependent on the soil moisture state of the model. This could account for features such as variable contributing areas, where a larger part of the incoming rainfall contributes directly to the runoff when the catchment is very wet. However, further analysis is required to establish this.

4.4.3 Example 3 – Application of combined framework

The approaches used in application examples 1 and 2 are combined here to form part of the framework introduced in Section 4.3.3. A larger number of structures than used in the other examples are also included here.

4.4.3.1 Data

The river selected for this study is again the Lower Medway at Teston (1256.1km^2) located in South Eastern England; see Section 4.4.2.1 for details.

4.4.3.2 Model structures

As stated earlier, a large variety of lumped parsimonious model structures can be found in the literature (e.g., Singh, 1995). However, the range of components on which these structures are based is relatively small. Some of the most commonly found components are selected here in a component library shown in Fig. 4.25. A detailed description of these components can be found in Chapter 3.

The soil moisture accounting components used are:

- The catchment moisture deficit (cmd, Evans and Jakeman, 1998; Kokkonen and Jakeman, 2001). A conceptual bucket with a bottom outlet to sustain drainage into the summer periods.

- The catchment wetness index (cwi, Jakeman and Hornberger, 1993). A metric approach related to the Antecedent Precipitation Index (API, e.g., Shaw, 1994).
- The probability distributed soil moisture stores (pd3 and pd4, Moore, 1999). A probability distribution of conceptual buckets based on a Pareto distribution. Evapotranspiration is either at the potential rate (pd3), as long as soil moisture is available, or at a rate declining linearly with soil moisture content (pd4).
- A simple bucket type structure (buc), evaporating at the potential rate as long as soil moisture is available.
- The Penman storage model (ic1, Penman, 1949). A two layered structure connected by an overflow mechanism. Evapotranspiration occurs at potential rate from the upper layer, similar to the root zone, and at a reduced rate, 1/12 of PE, from the bottom layer. An additional bypass mechanism diverts a fraction of the rainfall from the SMA component to contribute to the effective rainfall at time-steps where rainfall exceeds PE.

The routing components used are:

- Conceptual reservoirs in various combinations and in linear and non-linear form (e.g., Wittenberg, 1999).

4.4.3.3 Methodology

Multi-objective (MO) analysis and DYNIA are performed, based on the results of Monte Carlo sampling procedures. For the MO analysis, 20,000 parameter sets, *i.e.,* models, are randomly sampled from the feasible parameter space for each individual model structure, based on a uniform distribution.

For each of these models, five OFs are calculated: the overall RMSE and four OFs derived for different response modes of the catchment. The segmentation applied is based on an approach by Wagener and Wheater (2002) which uses the slope of the hydrograph and an additional threshold as segmentation criteria to split the hydrograph into different response modes (Fig. 4.26). The slope separates periods when the catchment is wetting up or is driven (Boyle *et al.*, 2000) by rainfall, *i.e.,* positive hydrograph slope, and when the catchment is draining, *i.e.,* falling

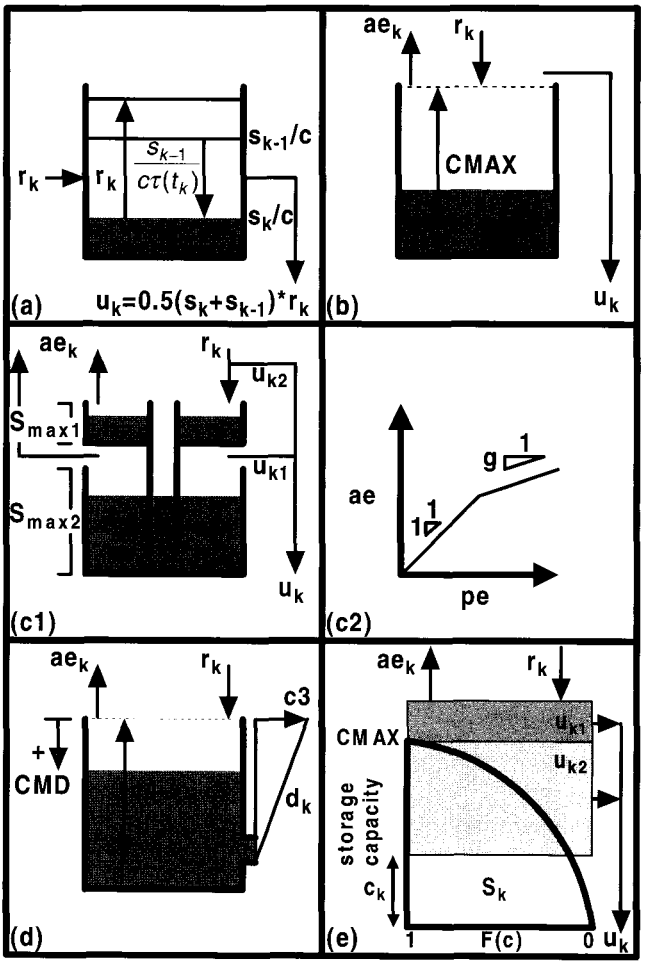


Fig. 4.25 Table showing the soil moisture accounting 'component library' used in the application example. The components are: (a) catchment wetness index (cwi), (b) simple bucket (buc), (c1) and (c2) Penman structure (ic1), (d) catchment moisture deficit (cmd), and probability distribution of soil moisture stores (pdX). (From Wagener *et al.*, 2002; reproduced with permission of the American Geophysical Union)

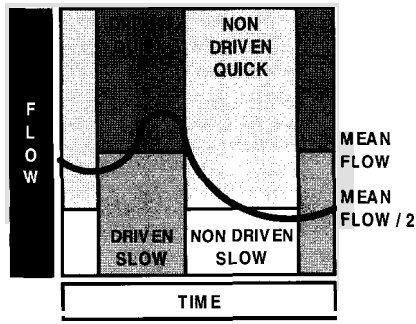


Fig. 4.26 Segmentation scheme used to derive multiple objective functions describing the model performance during different response modes.

hydrograph slope. A threshold is used to separate periods of high and low flow, *i.e.,* the mean flow during driven and 50% of the mean flow during drainage periods. Four OFs are therefore derived when the residuals during the different periods are aggregated separately using the RMSE criterion: FDH, driven flow during high flow; FDL, driven flow during low flow; FQ, quick drainage (high flows); and FS, slow drainage (low flows). This is a modification of the initial approach by Boyle *et al.* (2000), which was based on the analysis of flow and rainfall. However, the approach presented here has shown to be more suitable for British catchments as modelled in this example. These OFs are based on the assumption that different processes are dominant during periods of high and low flow, and during periods of catchment wetting-up and drainage.

The residuals, *i.e.,* the differences between observed and simulated flows, are calculated and summarized in form of the RMSE (see Equation 4.1) for each period. The performance and identifiability analysis is based on these measures.

The resulting parameter populations are used to rank all models or model structures, with respect to their performance and identifiability, using the measure introduced earlier. The best model structures are retained and a more thorough analysis using DYNIA is performed. DYNIA is based on a random sampling procedures using 2500 parameter sets collected from a uniform distribution. The smaller sample size is due to computational limitations of the current DYNIA application in the MATLAB (Mathworks, 1996) environment.

4.4.3.4 Results and discussion

The main results of the MO analysis as shown in Fig. 4.27 are as follows:

- At a general level for the SMA modules (Fig. 4.27, top) the probability distributions of storage elements (pd3 and pd4) seem to perform best, followed by the simple bucket (buc), and the cwi and cmd modules.
- The cm1 (*i.e.,* a cmd that always evaporates at the potential rate) performs much worse than the rest with respect to OFs that mainly describe periods of high flow, RMSE(total), FDH and FQ. This is also the case for the cmd module, but is not as pronounced. However, the cmd and cm1 modules do very well during low flow periods. This is caused by the bottom outlet of the bucket, which sustains the production of effective rainfall even during periods of severe moisture deficits in the SMA module.
- The overall result of the performance analysis is that the pd3 and pd4 SMA modules in combination with 2pll or 2pln routing modules are superior. The cmd is a useful component when the modelling purpose demands accurate prediction of low flow periods and periods of high flows are of minor importance.
- A detailed analysis of the routing components shows that the use of a non-linear conceptual reservoir in parallel with a linear one (2pln), performs better at the peaks (RMSE(total) and FDH), see Fig. 4.27(top).
- The uncertainty analysis (Fig. 4.27, bottom) however, reveals that the identifiability of the cmd parameters is very low and the module is rejected here on this basis. For some applications, this aspect might be of minor importance.

The pd3 and the pd4 SMA components are retained for further analysis with the DYNIA approach. Assuming that our interest lies in low flows, e.g., for water resources purposes, only a linear parallel routing structure (2pll) is considered. A non-linear component would be advisable for high flow periods.

The results of the DYNIA are shown in Fig. 4.28 and 4.29 for the structures pd3-2pll and pd4-2pll. The results reveal some problems with the pd3 SMA module.

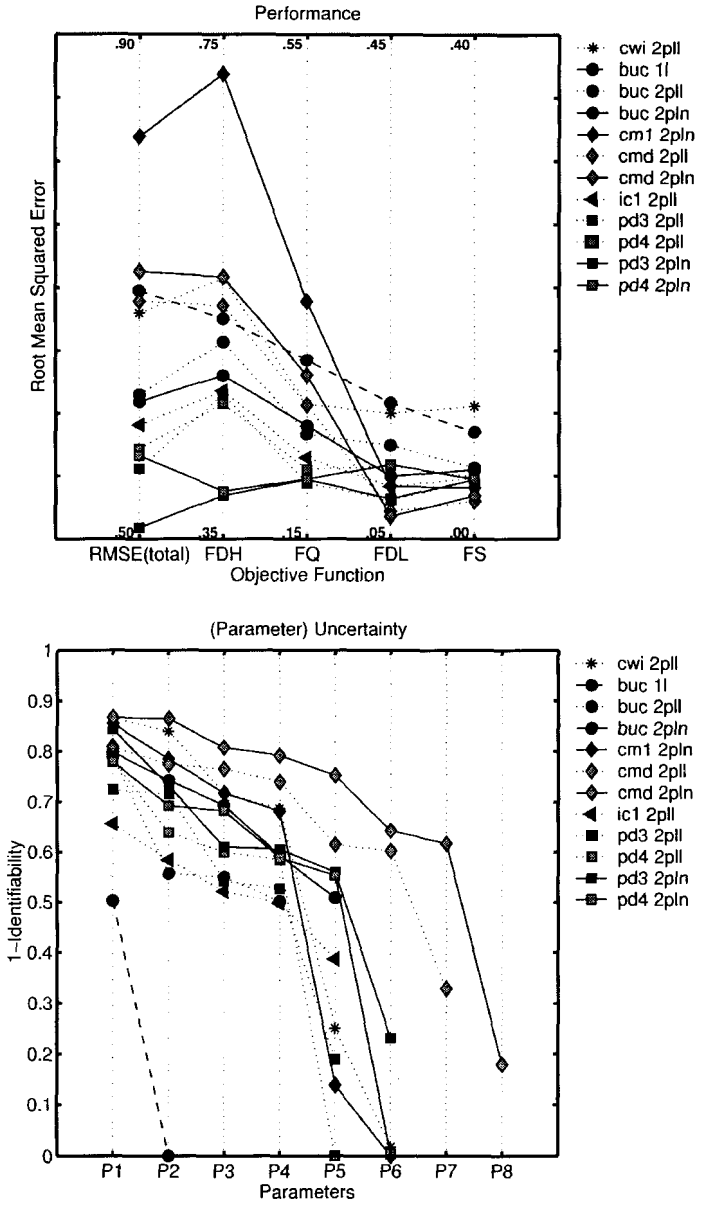


Fig. 4.27 Results of the model structure comparison. (From Wagener *et al.*, 2003b; reproduced with permission of the American Geophysical Union)

Figure 4.28 shows the dynamic identifiability of the five parameters of the pd3-2pll structure: (1) cmax, the maximum storage capacity; (2) b, the shape parameter of the Pareto distribution of storage capacities; (3) k(quick), the residence time of the quick linear reservoir; (4) alpha, the fraction of flow going through the quick flow component; and (5) k(slow), the residence time of the slow flow linear reservoir.

The plot for the parameter cmax exposes some ambiguity about its optimum value. The confidence limits (cfls) narrow into two different parts of the parameter space, towards low values after wet periods and towards high values during periods of wetting up, indicating structural inadequacies within the model structure. Similarly (but much less pronounced), the parameter b shows a slight shift of optimum after the wet period; the lower cfls go up. This is mainly identifiable during low flow events (e.g., dark areas just before time step 700). Figure 4.29 shows that the pd4-2pll structure does not have these problems.

The residence times of the routing component show the expected behaviour, *i.e.,* the cfls of k(quick) narrow down on the quick falling limbs of the hydrograph, while darker areas appear for k(slow) during the long recessions. The cfls for k(slow) hardly narrow during periods of identifiability, suggesting that the peaks on the response surface are quite small, and that the difference between values for this parameter is not large. Values for this parameter are therefore still widespread, since only the top 10% are selected here. The example of the two residence times also demonstrates the need for different window sizes. A small size (11 time steps) is required for k(quick), which has only very local influence, while a much larger window (81 time steps) is needed to capture the effect of k(slow). Finally, the parameter alpha is most identifiable during periods where the split between quick and slow response is occurring. However, further investigations, which are outside the scope of this example, are required to explain the behaviour of this parameter. In general, this structure is too simplistic to reproduce all aspects of the hydrograph with one parameter set. This is especially reflected in the results for cmax.

The difference between pd3 and pd4 is that, while pd3 always evaporates at the potential rate, pd4 decreases the evapotranspiration with decreasing soil moisture content in a linear manner. However, the actual

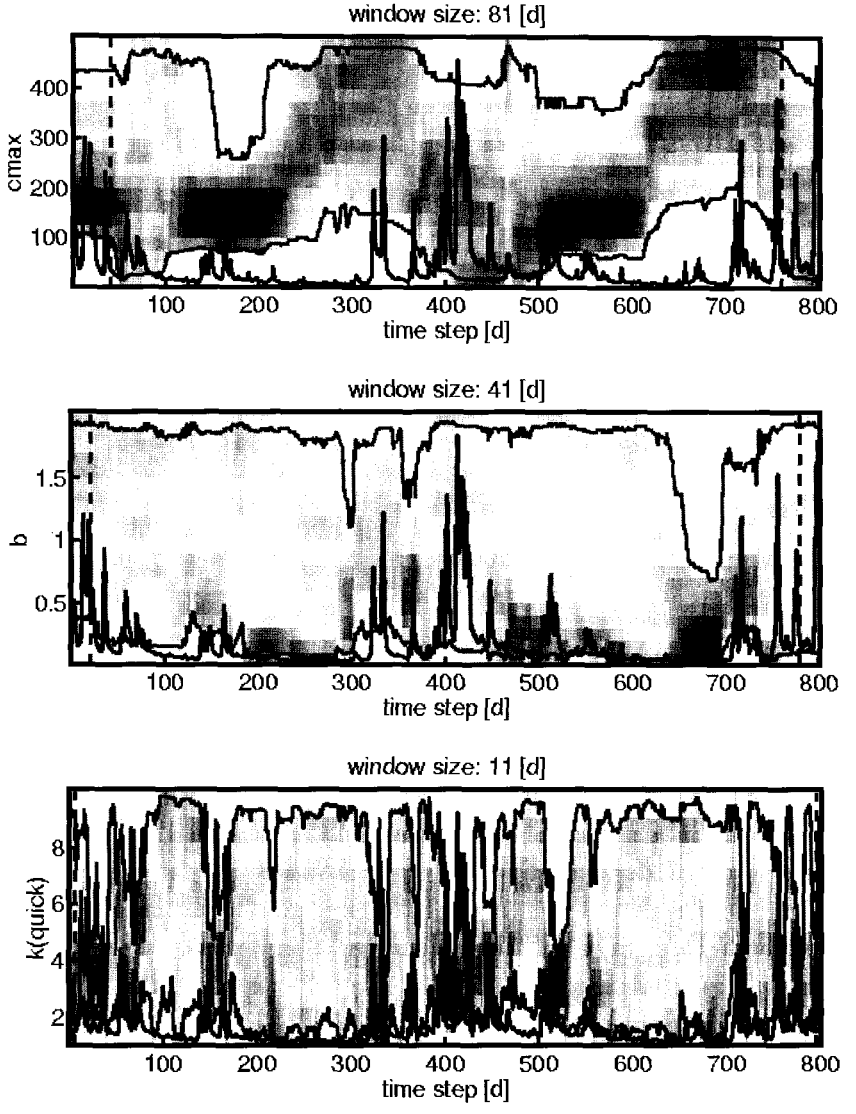


Fig. 4.28 DYNIA results for pd3-2pll. (From Wagener *et al.*, 2003b; reproduced with permission of the American Geophysical Union)

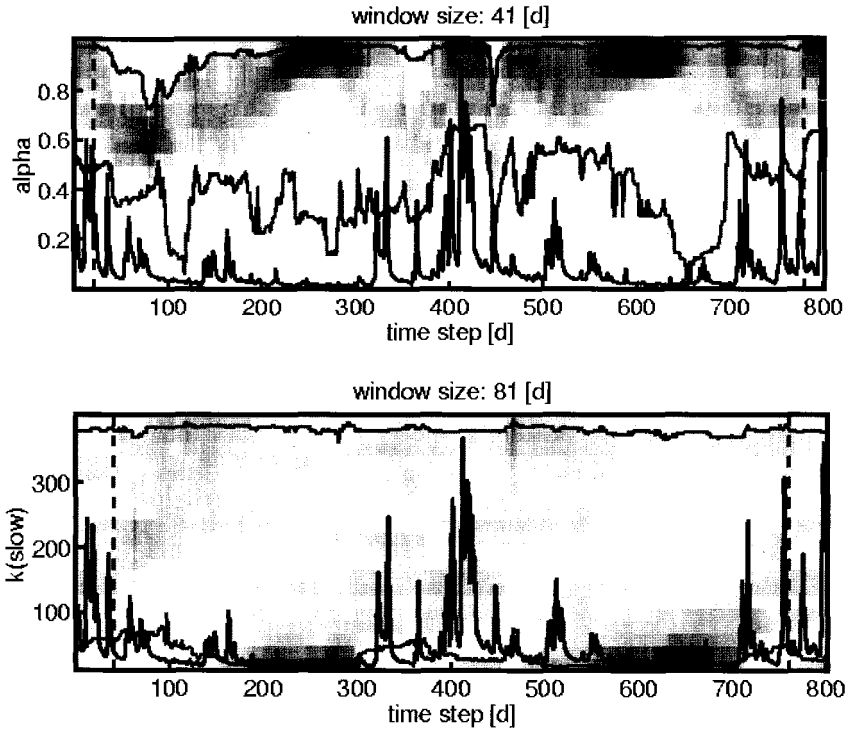


Fig. 4.28 (cont.)

evapotranspiration rate, *ae*, is related to the moisture content, *s*, in the store distribution without adding an additional (scaling) parameter, *i.e.*,

$$ae_k = \frac{s_k}{s_{\max}} \cdot pe_k \tag{4.3}$$

where $s_{\max}$ is the maximum soil moisture content and *pe* is the potential evapotranspiration, while *k* is a time step index.

The effect of this change can be seen in the dynamic results shown in Fig. 4.29. The ambiguity with respect to cmax is removed and the cfls only narrow towards larger values indicating a better structure. Also, the shape parameter b is much better defined.

It is interesting to remember that the MO performance analysis had shown that the pd3 component actually performed better. The reason is that the pd4 component puts an additional constraint on the behaviour of

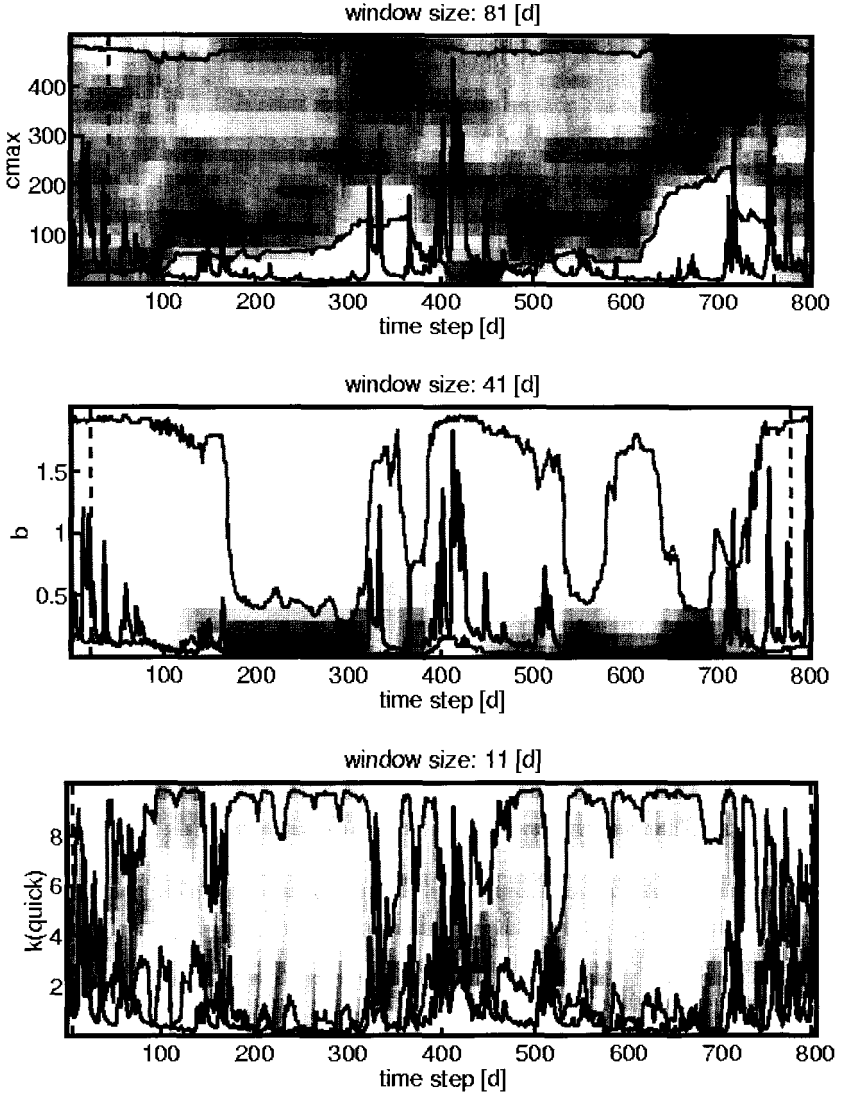


Fig. 4.29 DYNIA results for pd4-2pll. (From Wagener *et al.*, 2003b; reproduced with permission of the American Geophysical Union)

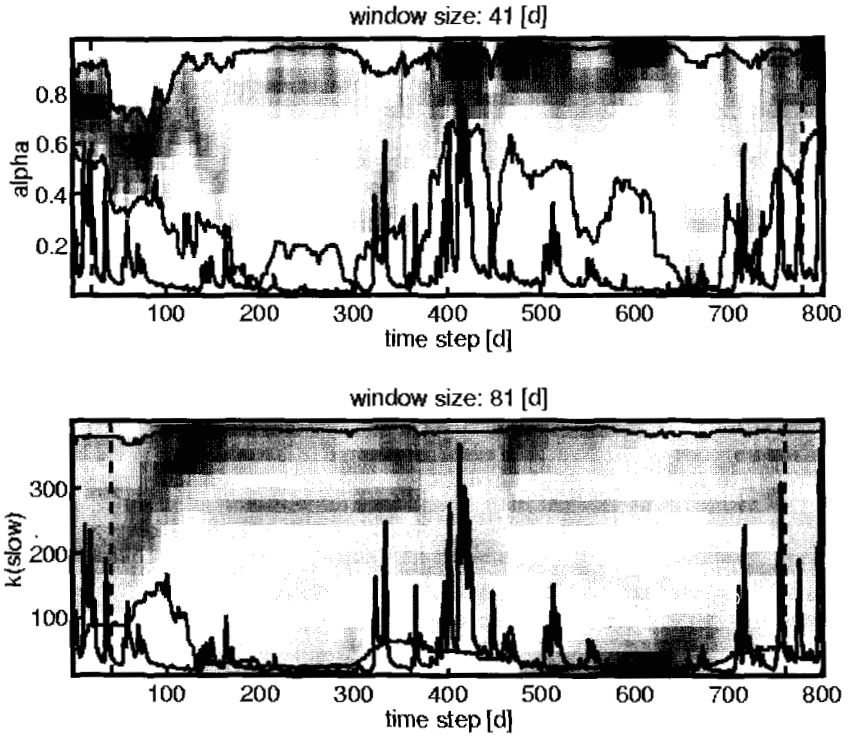


Fig. 4.29 (cont.)

the catchment system. This results in a less flexible model structure. The pd3 component can therefore perform better with respect to the different OFs. However, this is at the expense of a larger variation in parameter values as shown in the dynamic analysis. This indicates that pd4 is actually the better SMA component and should be retained, while pd3 should be rejected.

This result suggests that consistency in a model might be more important than optimality.

4.5 Summary and Conclusions

Current concepts underlying local CRR model identification, *i.e.,* the modelling of gauged catchments, have been analysed in detail. It has been suggested that the idea of establishing a *true* model structure is ill-founded and that it should be replaced by an evaluation of model structures with respect to

- performance,
- (parametric) uncertainty and
- underlying assumptions (reflecting for example structural inadequacy).

Additionally, current automatic methods of parameter identification lack the complexity of subjective manual techniques and are often too simplistic in their approach.

A multi-objective approach to model structure analysis has subsequently been developed to address aspects one and two of the three dimensional evaluation described above. Point three is assessed through a novel dynamic identifiability approach (DYNIA) which analyses how (initially uniform) parameter distributions are conditioned over different data periods. A violation of underlying assumptions is found, for example, if peaks in the response surface of a particular parameter vary in time and in (parameter) space. These approaches are combined in a framework of corroboration and rejection more in line with scientific methods – as propagated by Popper (2000) – than traditional procedures for rainfall-runoff modelling.

The following main observations were made in the application examples in Chapter 4:

- Example 1 led to the conclusions that accepting the multi-objective nature of model calibration and integrating it into the modelling process increases the amount of information retrieved from the model residuals to (1) find the parameter population necessary to fit all aspects of the observed output time-series (albeit separately), (2) increase the identifiability of the model parameters, and (3) assess the suitability of the model structure to represent the natural system (*i.e.,* identify model structural insufficiencies).

- One approach to derive multiple-objectives for single output models is the segmentation of the hydrograph. A segmentation based on the slope of the observed streamflow and additional thresholds has shown to be effective in separating different response modes for catchments in the UK (Chapter 4, examples 1 and 3), while an approach using rainfall to find 'driven' periods is difficult to apply due to the high frequency of rainfall events (see example in Chapter 3).
- A very distinct difference in optimum parameter sets to fit high and low flows has been found for all structures analysed in Chapter 4. This seems to be the main problem in currently available CRR model structures.
- The identifiability measure introduced in Chapter 3 is a very useful empirical tool to assess individual parameters in an objective manner without having to make limiting assumptions (examples 1 and 3).
- Applying this identifiability measure in a dynamic fashion leads to the DYNIA approach, which is of general utility and simple to implement. It can be applied to: (1) analyse individual model structures with respect to structural inadequacies; (2) identify suitable parameters in an approach more in line with sophisticated manual techniques; (3) separate periods of high and low information content for individual parameters, for example as a first step in a multi-objective analysis; (4) relate model parameters, and therefore model components, and response modes of the natural system; (5) investigate data outliers or anomalies; and (6) analyse the potential of experimental design, e.g., tracer experiments, to identify individual parameters.

Some of the main results of the DYNIA applications in examples two and three were: (1) For the Penman SMA component (ic1): (a) the root constant is mainly identifiable during wetting up periods; (b) the bypass parameter is defined by major rainfall events during dry periods; (c) the optimum bypass values change in time, indicating that a dynamic aspect is missing in the model structure; and (d) the correlation between root constant and bypass is reduced considerably if periods of high identifi-

ability only are considered. (2) For two parallel linear routing components (2pll): (a) the quick and slow flow residence times are mainly identifiable during periods of quick and slow recession; and (b) the split parameter alpha is most identifiable during the change between quick and slow flow dominated periods after large runoff events. (3) The probability distributed SMA (pd3 and pd4): (a) pd3 (ae at potential rate) performs better than pd4 (ae linearly dependent on soil moisture content) in a MO analysis; (b) however, the pd4 module shows more consistent optimum parameter values in a DYNIA, indicating that it is the better model structure, *i.e.,* in line with the underlying assumptions.

Chapter 5

Modelling Ungauged Catchments – Regional Procedures

The art of being wise is the art of knowing what to overlook.
William James (1842–1910)

5.1 Introduction

Rainfall-runoff models are commonly used tools to extrapolate streamflow time-series in time and space: in time, for example, to extend available streamflow records or predict the behaviour of catchments for different climate scenarios, and in space, to predict the response of catchments for which no (or only very short) time-series of streamflow measurements are available. The latter has been achieved with some success in terms of predicting the response of a catchment to an individual rainfall event, using (parametrically) simple models (e.g., NERC, 1975). Until recently the estimation of continuous streamflow time-series has remained an unsolved problem. However, new methods of analysis of model structure and parameter uncertainty allow the trade-off between performance and identifiability to be examined. By focusing on model structures associated with high identifiability, the possibility of relating model parameters to catchment characteristics has improved.

This chapter develops the review of Chapter 2 further. Chapter 5 is based on the assumption that many if not most RR model structures currently used for continuous modelling can be classified as conceptual, if the classification is based on two criteria (Wheater *et al.*, 1993): (1) the structure of the models is specified prior to any modelling being under-

taken, and (2) at least some of the model parameters do not have a direct physical interpretation in the sense of being independently measurable, and have to be estimated through calibration against observed data. Calibration is a process of parameter adjustment (automatic or manual), until catchment and model behaviour show a sufficiently high degree of similarity to satisfy the hydrologist. The similarity is usually judged by one or more objective functions (OFs) accompanied by visual inspection of observed and calculated hydrographs.

However, as discussed in Chapter 2, the need for calibration is problematic if the modelling task requires extrapolation of flow predictions to a location for which no or only short observed streamflow time-series are available. Unfortunately this is often the case: even in the UK where about 1400 gauging stations are in operation, many small catchments remain ungauged (Sefton and Howarth,1998). Alternative approaches to derive parameter values are required under those circumstances. Early attempts have simply used the parameter values derived for neighbouring catchments where streamflow data are available (a geographical proximity approach; e.g., Mosley, 1981; Vandewiele and Elias, 1995). However, this seems to be insufficient since even nearby catchments can be very different with respect to their hydrological behaviour (Post *et al.*, 1998; Beven, 2000b).

Some authors propose the use of parameter estimates directly derived from: soil properties such as porosity, field capacity and wilting points (to derive model storage capacity parameters); percentage forest cover (evapotranspiration parameters); or hydraulic conductivities and channel densities (time constants) (e.g., Koren *et al.*, 2001; Atkinson, 2001). The measurements of soil properties are usually made on a small (laboratory) scale; this approach assumes that it is possible to derive some *effective* catchment-wide values using simple aggregation rules, e.g., average soil properties when different soil types are present in the catchment. The approach does not consider the effects of larger scale heterogeneity, for example due to preferential flowpaths by macropores, or the potential scale-dependence of effective parameters. It is then assumed that the effective soil values are equivalent to conceptual model parameters. Conceptual model parameters describe an integrated catchment response usually aggregating significant heterogeneity (and including the effect of

preferential flow paths). It seems unlikely that they can be derived easily from soil properties that do not consider all influences on the flow of water through the catchment. However, there is some suggestion that one might gain reasonable initial parameter ranges (or estimates) that can be used as a starting point in a calibration procedure (Duan *et al.*, 2001; Koren *et al.*, 2001). Further fine tuning of these estimates using locally observed flow data is however needed "because the physical information available to estimate a priori parameters is not adequate to define local physical properties of individual basins for accurate hydrologic forecasts" (Duan *et al.*, 2001). The authors are not aware of any large scale study so far that compares locally calibrated and *measured* parameters for a large number of catchments to statistically corroborate or reject this approach.

This type of approach is unlikely to yield satisfactory results for the simple parsimonious structures applied in this monograph. The most commonly used approach to ungauged modelling, using this model type, is to relate model parameters and catchment characteristics in a statistical manner (e.g., Jakeman *et al.*, 1992; Sefton *et al.*, 1995; Post *et al.*, 1998; Sefton and Howarth, 1998), assuming that the uniqueness of each catchment can be captured in its unique combination of characteristics. The basic methodology is to calibrate a specific model structure, here called the *local model structure*, to as many catchments as possible and derive statistical (regression) relationships between the (local) model parameters and the catchment characteristics. This statistical relationship, here called the *regional model*, and the measurable properties of the ungauged catchment can then be used to derive estimates of the (local) model parameters (see Fig. 5.1). This procedure is usually referred to as *regionalization* or *spatial generalization* (see, for example, Lamb and Calver, 2002).

It is evident from the literature review (Chapter 2) that few studies have been performed to analyse this type of regional modelling in detail. While local modelling (particularly the aspect of model calibration) has been at the centre of research for decades, continuous regional modelling has only recently become a focus (e.g., Post and Jakeman, 1996; Sefton and Howarth, 1998; Seibert, 1999b; Lamb *et al.*, 2000). The main reason for this new interest is the hope that a parsimonious model structure, in

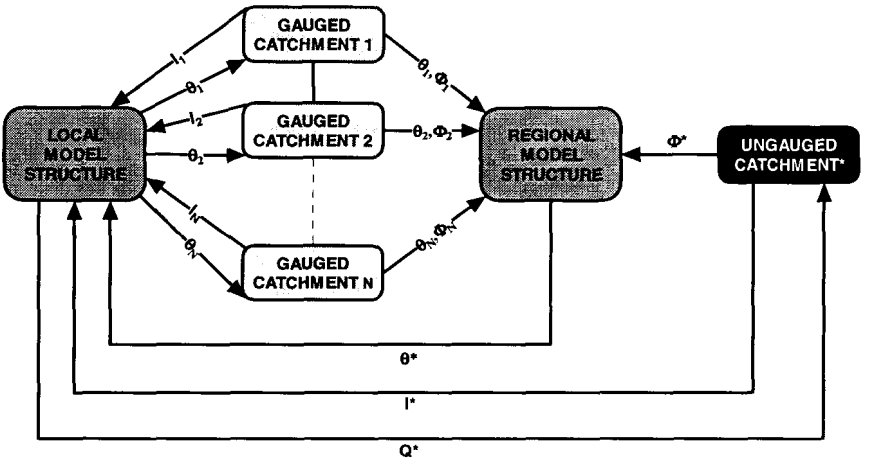


Fig. 5.1 The idea of conceptual parameter regionalization. Where Φ are catchment characteristics, Q is the streamflow, θ are model parameters and I are input time-series.

combination with improved methods of optimization and identifiability analysis, could yield satisfactory results. A range of parsimonious structures with identifiable parameters is available today (see Chapter 3), while older model structures often contain an excess number of parameters which makes the identification of a unique parameter set very difficult at best (e.g., Johnston and Pilgrim, 1976). However, the regionalization studies that can be found in the literature are mainly relatively straightforward applications without a detailed analysis of the components and uncertainties involved (see Chapter 2 for details). Additionally, the large variety of catchment characteristics used and differences in climate make it very difficult to compare the results of these studies.

In this chapter both theoretical and applied issues of the regionalization of continuous parsimonious CRR models are considered. The first section reviews the idea of regionalization and the elements involved in a typical procedure. Alternative components for regionalization are discussed and a regional modelling framework is suggested. The second section provides an application example where the suggested framework is applied to ten catchments in southeast England. The aim of this appli-

cation is not to provide a comprehensive case study, but rather to test different possible components within the framework for their suitability.

5.2 The Idea of Regionalization

The objective of each rainfall-runoff modelling exercise is the calculation of runoff from a catchment as a function of rainfall and various meteorological forcing variables, such as temperature and relative humidity. This can be represented mathematically as follows

$$Q = M_L(\theta_L \mid I) + \varepsilon_L \tag{5.1}$$

where Q is the simulated streamflow; I is a matrix of input variables (e.g., rainfall and temperature); M_L is a given (local) model structure, θ_L is a vector of parameters within this structure; and ε_L is an error term. The model parameters will usually be estimated through a calibration exercise in cases where measured time-series of runoff over a sufficiently long period are available. The required length of the time-series depends, among other things, on the complexity of the model structure used and the information content of the data. It might range from three years for a simple structure (Sefton and Howarth, 1998), to up to a decade for more complex model structures (Yapo *et al.*, 1996). However, in principle the data set should always be long enough to avoid the problem of the parameters being representative only of a particular climate sequence (Gan and Burges, 1990).

If no runoff data are available for a specific catchment, *i.e.,* if it is ungauged, an attempt can be made to calibrate the model structure to a large number of gauged catchments and to find a functional relationship between the usually individual conceptual model parameters (dependent variables) and the catchment characteristics (independent variables), *i.e.,* a *regional* model structure of the following type:

$$\hat{\theta}_L = M_R(\theta_R \mid \Phi) + \varepsilon_R \tag{5.2}$$

where $\hat{\theta}_L$ is the estimated model parameter at the ungauged site; $M_R(\cdot)$ is a functional relation for $\hat{\theta}_L$ using a set of physiographical and meteorological catchment characteristics Φ, while θ_R is a set of regional model parameters; and ε_R is an error term. The functional relationship is

called the *regional model structure* in the context of this paper. One model, *i.e.*, (regional) model structure and (regional) parameter combination is normally derived for each (local) parameter. This means the model parameters are assumed to be independent.

No generally accepted procedure for regionalization of conceptual, continuous models currently exists. However, the following steps are typically found and therefore given here as the basic outline of a regionalization procedure (see also Fig. 5.1):

(1) Decide which (sub-)set of catchments can be described by a single local model structure M_L and a single regional model, *i.e.*, one structure M_R with a specific set of parameters θ_R. Catchments, which are too different with respect to their dominant hydrological processes, might require different local model structures to represent them in a physically (or conceptually) realistic manner.

(2) Collect a set of catchment characteristics Φ related to the hydrological response of the selected catchments. Examples are size, shape, drainage density, land use, and soil properties.

(3) Apply the local model structure M_L to each of the gauged catchments and estimate the *optimum* parameter set (or population) θ_L for each catchment.

(4) Relate the derived individual parameter values θ_{Li} and the catchment characteristics Φ using the regional model structure M_{Ri}. (Different regional model structures might be necessary for different local model parameters).

(5) Apply the regional model $M_{Ri}(\theta_{Ri} \mid \Phi)$ to estimate each parameter $\hat{\theta}_{Li}$ for the ungauged catchment.

(6) Predict flow in the ungauged catchment using parameter set $\hat{\theta}_L$.

There are of course variations on this approach, for example the recursive incorporation of information from step 4 in 3 (e.g., Vogel *et al.*, 1999; Lamb and Calver, 2002). This means that the objective is not just optimal local performance, but also an optimization of the performance of the regression relationship. However, this listing contains all of the relevant components. These components and aspects of regionalization are discussed in the following section.

5.3 Problems and Possible Ways Forward

Unfortunately, each of the steps outlined above contains uncertainties that are unavoidably propagated into the regionalization result (such as the streamflow prediction at an ungauged site), even if those uncertainties are often ignored in regionalization studies. One can split the uncertainties or problems related to the individual components into two groups: 1) the problems related to local modelling, *i.e.,* those related to the selection and calibration of the local model structure to every individual catchment, and 2) additional problems due to the need for spatial extrapolation using a regional model. The main problems are:

- Selection of *catchment properties*. What are suitable characteristics to describe and cluster (pool) catchments with respect to their hydrological response? This is an important aspect with respect to both the local and the regional modelling steps.
- Selection of the *local model structure*. Determine potential inadequacies and lack of identifiability with respect to the conceptual structure used.
- Identification of the *local parameters*.
- Selection and identification of the *regional model structure and its parameters*. What is the nature of the relationship between catchment characteristics and model parameters?
- Selection of the *regionalization procedure*. For example, is the calibration-objective purely the optimization of the performance of the local model in each catchment, or is the performance of the regional model considered?

The problems related to local modelling are only discussed briefly, since these aspects were treated thoroughly in Chapter 4. The regionalization or regional modelling problems are analysed in detail.

5.3.1 Catchments, their characteristics and clusters

A large number of catchments with good quality data and characteristics are required to derive statistically significant relationships in a regionalization study (e.g., Post *et al.*, 1998).

Since there is no established theory to relate catchment characteristics and the parameters of conceptual model structures, a trial and error approach is normally adopted in which a wide range of available characteristics is considered initially and then reduced, based on statistical analysis and hydrological reasoning (including results of earlier studies). The correlation between catchment characteristics can be analysed, leading to rejection of some characteristics (because they contain the same information as others) or the combination of characteristics to form new variables using principal component analysis (Howarth, 1998). However, in the latter approach there can be difficulties in the loss of a unique physical association with a given characteristic leading to further difficulties in explaining the correlation results. This option is therefore not considered further here.

The catchment characteristics are also very important with respect to the clustering or pooling of catchments. It is usually assumed that a *general model structure* exists that is capable of representing all catchments included in the study. Catchments that are too extreme are excluded (e.g., Sefton and Howarth, 1998), while the remaining variability is captured by the individual parameter sets. However, a more general approach would include a large variety of catchments and identify the dominant variable(s) (catchment characteristics) that allow the allocation of a particular catchment to a particular (local) model structure. Since the idea is to allow selection of different regional models for different clusters of catchments regardless of geographical location, the term regionalization is slightly misleading since it implies that only a single geographical region is considered.

For UK flood design, Robson and Reed (1999) suggest pooling (clustering) catchments for flood frequency analysis based on their hydrological similarity. This supercedes an earlier approach based on geographic regions (NERC, 1975). Catchments within a region can show considerable heterogeneity and the earlier approach therefore showed problems of robustness (Kilsby and O'Connell, 2001). The new methodology suggests description of similarity using three features: size, wetness and soil properties, represented by the catchment size, AREA [L^2], the Standard-period Average Annual Rainfall, SAAR [L], and the Base Flow Index from the Hydrology of Soil Types Classification, BFIHOST [-] (see dis-

cussion below), respectively. It can be noted that the baseflow index defines the proportion of hydrograph classified as baseflow, and is thus an important descriptor of the hydrograph response which reflects the geological controls on catchment response. The similarity between catchments is based on the region-of-influence approach introduced by Burn (1990), and is measured using the Euclidean distance ($dist_{ij}$) in the three-dimensional space,

$$dist_{ij} = \left[\frac{1}{2} \left(\frac{\ln AREA_i - \ln AREA_j}{\sigma(\ln AREA)} \right)^2 + \left(\frac{\ln SAAR_i - \ln SAAR_j}{\sigma(\ln SAAR)} \right)^2 + \left(\frac{BFIHOST_i - BFIHOST_j}{\sigma(BFIHOST)} \right)^2 \right]^{\frac{1}{2}} \tag{5.4}$$

where σ is the standard deviation of a variable. The logarithmic transformation ln leads to more symmetrical distributions and to the fact that the distance measure (for AREA and SAAR) is based on ratios, rather than differences. Halving the weight of AREA increases the influence of the other two variables. The computed value of the measure has been found to be approximately 0.5 within a single pooling group (Robson and Reed, 1999, p. 158).

For the purpose of identifying groups for which a particular model structure is suitable, a simpler approach than that of Robson and Reed might suffice. Various studies have shown that a single model structure can be suitable for a wide range of different catchment sizes (e.g., Littlewood and Jakeman, 1992; Jakeman and Hornberger, 1993). Also, the climatic variable (SAAR) might be less important for model choice than a variable describing soil types and geology of the catchment (such as BFIHOST) if all catchments are located in the same climatic region. Boxall (2001) and Lee (2001) for example found that BFIHOST was a dominant characteristic in explaining the hydrological variability of catchments in the southeast of England, which suggests that a clustering (pooling) based on this variable is a suitable approach for the purpose of

local model structure selection in this geographic region. However, other variables might be more appropriate for other groups of catchments and a wider range of characteristics should in general be investigated. The idea is to derive different regional models for different clusters of catchments, regardless of the geographical location of the catchments included. The term regionalization is therefore slightly misleading since it implies that only a single geographical region is considered. Other variables might be more appropriate for other groups of catchments and a wider range of characteristics should be investigated.

It is useful to describe the HOST system (Boorman *et al.*, 1995) in slightly more detail. HOST is a hydrologically based classification of UK soils using a number of conceptualizations of processes occurring in soil (or substrate). The conceptualizations are based on three basic settings: soil over permeable substrate with shallow or deep groundwater, and soil over shallow slowly permeable substrate. Adding soil properties from soil profile descriptions (e.g., a peaty top layer) and wetness regimes (e.g., as indicated by the presence of gleying) leads to eleven models. Further considering other soil properties and the geology of the substrate derives twenty-nine HOST classes. The utility of the HOST approach was validated through estimating measured Base Flow Index (BFI) values, *i.e.*, a catchment scale hydrological variable, from a regression on the HOST classes. The regression yielded a coefficient of determination (Eq. 5.26) of 0.79. An example calculation is given in the application at the end of this chapter.

There is also the question of how well the available catchment characteristics can describe the hydrological behaviour of a catchment in order to derive regional models for all model parameters. In humid regions, the behaviour will be dominated by subsurface characteristics, while most catchment characteristics available refer to the surface. "Essentially there is much about the subsurface flow domain that will remain unknowable, a fortiori, and is revealed only in the response observed at larger scale, with all the limitations of knowledge that these observations imply" (Beven, 2000b).

The variable BFIHOST will be used in this study to cluster catchments, since UK data are utilized in the application example. No other variables will be investigated further here. One has to estimate then the

range of local model structures required to represent the different clusters. Further research is required to devise more generally applicable schemes. This was not possible here due to the focus on UK data.

5.3.2 Local model

A model structure capable of representing the response of all catchments within a particular cluster must be found. The basic assumption underlying current regionalization approaches is that only a relatively simple model structure with identifiable parameters offers a chance to develop sensible relationships between its parameters and catchment characteristics (Wheater *et al.*, 1993).

The problem of lack of identifiability of model parameters is one of the major problems in CRR modelling today and is well documented in the research literature (e.g., Beven and Binley, 1992; Wagener *et al.*, 1999; 2001). The bottom line is that there is often an ambiguity about which parameter values are representative of the catchment under consideration. The result of this is that the statistical regression relationships (regional models) derived are virtually meaningless if equally acceptable parameters are widely distributed over their feasible ranges. In most regionalization studies (e.g., Sefton and Howarth, 1998), only the best value of each parameter, with respect to a specific OF, is selected to create relationships with catchment characteristics. However, this selection can be rather arbitrary due to a lack of parameter identifiability.

This is one reason why simple model structures, which only concentrate on the main aspects of the hydrological response of a catchment, are often used for regionalization. These models contain a small number of parameters, which are usually identifiable. Statistical relationships to catchment characteristics can therefore be established with more confidence. However, the structure is also less flexible and fewer processes are described. Examples are event-based models like Unit Hydrograph approaches (e.g., Nash, 1960; NERC, 1975; Tung et al., 1997), models that predict only low flows (e.g., Chang and Boyer, 1977; Nathan and McMahon, 1992), or continuous models working on large (e.g., monthly) time-scales, Vandewiele *et al.*, 1991; Vandewiele and Elias, 1995).

The problem increases when continuous models, working with a relatively small discretization in time (daily or less) are to be regionalized. These models usually have between five and fifteen parameters and therefore often severe problems of non-uniqueness of parameters. As noted earlier, it seems that only between three and five parameters are identifiable using a single OF (Gupta, 2001). However, new approaches of multi-objective (e.g., Boyle *et al.*, 2000; Wagener *et al.*, 2001) and dynamic (Wagener *et al.*, 2002b; 2003a) parameter estimation might reduce this problem (see Chapter 4 for details). The quality of predictions made with these simple structures is also often equal or superior to that of more complex model structures (see Section 2.4).

Typical representatives of parsimonious and relatively general model structures, capable of continuous modelling at a daily (or even sub-daily) time-step that are already used for regionalization are the PDM (Calver *et al.*, 1999; Lamb and Calver, 2002), the IHACRES (Jakeman *et al.*, 1992; Sefton and Boorman, 1995; Sefton *et al.*, 1997; Post *et al.*, 1998; Sefton and Howarth, 1998), the TATE (Calver *et al.*, 1999), and the abcd (Vogel *et al.*, 1999; Fernandez *et al.*, 2000) model structures (see Chapter 2 for details).

A schematic representation of a general parsimonious model structure is given in Fig. 5.2. The following structural components are usually considered: a component describing losses due to actual evapotranspiration (ACTE), a component describing storage of moisture in the catchment (STORE), and a parallel routing component, with one component describing the quick response (QUICK) and one describing the slow response (SLOW) of the catchment. However, this division is not a strict division into different pathways, since the quick response is commonly a combination of surface and sub-surface flows (e.g., Ward and Robinson, 2000, p.236). Commonly, a constant division between the contribution to the quick and slow response (SPLIT) is applied, while other model structures relate this division to the actual moisture content, as in some versions of the PDM model structure. A constant split has the advantage that it can sometimes be derived *a priori* from catchment characteristics such as the baseflow index (e.g., BFIHOST) or the standard percentage runoff (SPRHOST), as done for example by Young (2002) or Lamb and Calver (2002).

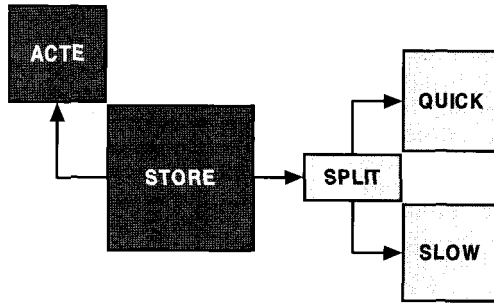


Fig. 5.2 Schematic description of a general model structure with components (parameters) describing the actual evapotranspiration (ACTE), the storage volume, (STORE), the split between quick and slow response (SPLIT), and the quick and slow response residence times.

Another problem is the influence of the choice of OF in combination with the problem of model structural inadequacies. Different parameter sets might be optimal depending on the choice of OF. Do they therefore represent different processes and are they therefore related to different catchment properties? As mentioned before, CRR model structures commonly aggregate (in space and time) the hydrological processes occurring in a catchment into a number of key responses, represented by storage components (state variables) and their interactions (fluxes). The model parameters describe aspects such as the size of those storage components, the location of outlets, the distribution of storages within the catchment, etc. Conceptual parameters, therefore, usually refer to a collection of aggregated processes and they "may cover a large number of sub-processes that cannot be represented separately or explicitly" (Van Straten and Keesman, 1991). This is sometimes demonstrated by the fact that optimization with respect to different OFs can result in different optimal values for a single parameter (Wagener *et al.*, 2001, and Chapter 4). It might therefore also lead to different regional models.

The choice of an OF can also influence the amount of correlation between the parameters. The usual assumption in the regional (regression) model is that all model parameters are uncorrelated. However, this assumption is unlikely to be justified in most CRR models. One reason for this is that it is very difficult to define an OF that singles out the influence of a single parameter, *i.e.*, a model component. For example, the

performance of different residence times measured using the distance between the observed and calculated hydrographs will depend on the volume under the hydrograph, which is dependent on the soil moisture accounting (SMA) model component.

Beven (2000b) mentions that certain parameters only have a meaning within a particular model structure, even if their name is identical to parameters in other model structures. Kokkonen and Jakeman (2001) tested this and found that identical components used in different versions of the IHACRES model structure can result in different optimal model parameters. They combined a parallel transfer function routing component alternatively with a parametric (similar to sma_cmd) and a metric (similar to sma_cwi) SMA component. They subsequently found that the routing parameters (*i.e.,* residence times and the constant split between quick and slow response), were a function of the SMA component selected. Use of the metric component resulted in higher residence times and a larger contribution to the slow response for 45 different calibration experiments. Kokkonen and Jakeman (2001) see the explanation in the extra delay introduced by the parametric component; the main part of the effective rainfall is produced when the (single) reservoir is overflowing, while the metric structure produces a more consistent output.

Not only is the appropriate parameter set hard to estimate, recent research has also shown that different model structures can sometimes be indistinguishable (e.g., Uhlenbrook *et al.*, 1999). This result is related to an increasing awareness of model structural inadequacies (e.g., Gupta *et al.*, 1998; Boyle *et al.* 2000; Wagener *et al.*, 2001; 2003a; b). An inadequate model structure will yield unsuitable parameter values and their relationship with catchment characteristics will, even if statistically successful, not be meaningful in any hydrological sense. A physical interpretation of these parameters, and therefore of the regional relationship, will be of little value. The requirement is, therefore, to derive a physically realistic representation of the hydrological system under study with as small a number of parameters as possible to achieve high parameter identifiability.

Several questions with respect to local models require investigating. What is the optimum model (or population of models or even model

structures) for a particular catchment, and therefore also cluster of catchments? The effect of different OFs on the local parameter estimates and on the subsequent regional model has to be examined. Another question to be asked is to what extent are local parameter estimates dependent on the choice of structures within a single model?

5.3.3 Regional model

A second model is required to relate the identified parameters to the catchment characteristics. This is usually a regression model of some type. Some possible models are described below.

5.3.3.1 Univariate regression

The most common approach adopted is the use of multiple regression model structures (e.g., Nash, 1960; NERC, 1975; Mimikou and Gordis, 1989; Nathan and McMahon, 1992; Burn and Boorman, 1992; Dyer *et al.*, 1994; Sefton and Howarth, 1998; Vogel *et al.*, 1999; etc.).

This structure can be written in matrix form (e.g., Weisberg, 1980, p.41) as

$$\theta_L = \Phi\theta_R + \varepsilon_R \tag{5.5}$$

where θ_L is a vector of the model parameter (the dependent variable or response), Φ is a matrix of measured catchment characteristics (the independent variables or predictors, usually containing a column of 1's to estimate the intercept), θ_R is a vector of regression parameters, and ε is an error vector. The ordinary least squares solution $\hat{\theta}_R$ of θ_R can be estimated using

$$\hat{\theta}_R = \left(\Phi^T\Phi\right)^{-1}\Phi^T\theta_L \tag{5.6}$$

(Weisberg, 1980, p.42) where Φ^T describes the transpose of the matrix Φ. However, most software packages use more sophisticated approaches because the least squares solution can have poor numeric properties (Mathworks, 1993).

Non-linearity is sometimes included through the transformation of the dependent or independent variables used in the regional model structure (e.g., Tung *et al.*, 1997).

However, univariate regression assumes that the dependent variables (the model parameters) are not correlated, an assumption that is often not justified, as outlined earlier.

5.3.3.2 'Correlated' regression

An attempt to consider parameter dependencies was made by Tung *et al.* (1997) using two approaches called multivariate regression analysis (MVR) and seemingly unrelated regression (SUR) for the regionalization of a two-parameter Nash Instantaneous Unit Hydrograph (NIUH) model structure.

Several correlated dependent variables can be considered simultaneously in MVR. However, MVR requires that all dependent variables (model parameters) be related to exactly the same independent variables using identical functional relationships.

A more general extension of the MVR is the SUR. Here the functional relationships for the different dependent variables can vary. Tung *et al.* (1997) suggest a generalized least squares approach to estimate the regression parameters while considering the correlation between the dependent variables.

They found that considering the correlation between the NIUH parameters when deriving the regional equation improved their predictive capability. However, appropriate software tools (including correlated regression approaches) were not available. Deriving tools to perform this type of regression for more complex model structures than the two-parameter NIUH used by Tung *et al.* (1997) is beyond the scope of this study.

Campbell and Bates (2001) used the Markov Chain Monte Carlo approach by Campbell *et al.* (1999) to derive local posterior parameter distributions for the two parameters of the RORB model structure (Laurenson and Mein, 1995) for 39 Australian catchments. They subsequently simultaneously regionalized the distribution means for the different parameters, assuming normality of the posterior distributions and

using a regression approach similar to the one of Tung *et al.* (1997). Campbell and Bates (2001) also included information about the local parameter covariance matrices into their regression, which reduced the standard error on the regional parameter estimates.

There is no reason, however, why the relationship between a particular model parameter and one or more catchment characteristics should be in a particular pre-supposed form. An approach that does not presuppose a particular functional relationship could be advantageous in exploring this. One such an approach is the Artificial Neural Network (ANN), discussed below.

5.3.3.3 Generalized regression – radial basis neural networks

An ANN mimics the way biological nervous systems function. It consists of simple elements that work in parallel and are determined by their connections.

A Generalized Regression Neural Network (GRNN) can be used for function approximation (Wasserman, 1993). It can approximate a continuous function to any level of accuracy if enough hidden neurons are available (Demuth and Beale, 1997, p.209). A GRNN consists of two layers, a radial basis layer and a second (special) linear layer. The input to the GRNN (in this case a matrix of catchment characteristics) is combined with weights and biases through linear filters. The GRNN is calibrated, *i.e., trained*, by adjusting weights and biases. The result provides the input to the hidden layer, the radial basis layer. The activation or transfer function for a radial basis neuron takes the following form

$$f(x) = e^{-x^2} \tag{5.7}$$

where x is the filtered input and f is the activation function (Demuth and Beale, 1997, p.197). A second layer of filters is then applied to the output of the hidden layer. An additional layer of activation functions follows this layer of filters. The activation functions are, in the case of a regression neural network, of linear form (Gershenfeld, 1999, p.152).

Major disadvantages of most ANNs are 1) the functional relationship cannot be stated explicitly and 2) it is difficult to incorporate uncertainty in the predictions. The first problem means that the ANN remains a black

box. Additionally, ANNs require a large number of data points to be trained, but those are often not available in regionalization studies. Another emerging technology, symbolic regression, could be used to overcome this problem.

5.3.3.4 Symbolic regression – genetic programming

Symbolic regression (SR) is a field of genetic programming (GP), which itself is a type of evolutionary algorithm. In SR, the aim is not just to identify the parameters of the functional form of the relationship between dependent and independent variable, but also the functional form itself (Koza, 1992, 2000). This function is derived together with the required parameters during the calibration process.

The steps in GP are similar to those in other evolutionary algorithms. GP is inspired by the Darwinian theory of evolution, in which the survival and breeding of the fittest members evolves a population. The algorithm starts with the random specification of an initial population, *i.e.,* mathematical elements (e.g., logarithmic transformation, subtraction or multiplication) and parameters: the initialization stage. Some members of this population are selected to produce offspring. The selection is usually related to the fitness of the members measured by an OF with better performing members having a higher chance of being selected. These members then breed, *i.e.,* new random combinations are formed. The procedure is usually terminated when at least one member of the population produces a sufficiently high performance.

There are few examples of the use of GP in hydrology (e.g., Drecourt, 1999; Babovic and Bojkow, 2001). GP is computationally demanding and its use in hydrology is still a research topic on its own. However, GP offers an interesting alternative to conventional regression and might be worthy of future exploration.

Whatever regional model structure is adopted, unavoidable uncertainties that arise, for example, from problems in parameter estimation must be considered and propagated into the flow prediction at the ungauged site. While standard confidence limits can be derived for the parameters (and therefore predictions) of a normal regression model, there are also complementary or alternative approaches such as weighted or fuzzy re-

gression. Only weighted regression is analysed in this section; other approaches to considering uncertainty are explained in section 5.3.5.

5.3.3.5 Univariate weighted regression

The identifiability of individual parameters might vary between different catchments, e.g., a particular model component might be more suitable for one catchment than another, even if the model structure is generally suitable for all. Some parameter estimates will therefore be more reliable than others. It seems sensible to give a higher importance to the more reliable parameter estimates when deriving a regional model. A measure of identifiability, based on the slope of the cumulative parameter distribution, was introduced in Chapter 4. This measure can be used as a weight in a regression model to enhance the influence of better-identified parameters on the final regional model. This can be written in mathematical form as weighted linear regression (e.g., Weisberg, 1980, p.75f.). Starting with the regression model introduced in Eq. 5.5,

$$\theta_L = \Phi\theta_R + \varepsilon_R \tag{5.8}$$

the generalized least squares solution is given (in contrast to Eq. 5.6) by

$$\hat{\theta}_R = \left(\Phi^T \Lambda^{-1} \Phi\right)^{-1} \Phi^T \Lambda^{-1} \theta_L \tag{5.9}$$

in case of a weighted least squares, where a higher weight indicates a more precise estimate. The weighting function Λ can be written as

$$\Lambda = \begin{bmatrix} w_1^{-1} & & & \\ & w_2^{-1} & & 0 \\ & & \ddots & \\ 0 & & & w_n^{-1} \end{bmatrix} \tag{5.10}$$

one can then define an $n \times n$ matrix, C, for which $C^T C = \Lambda^{-1}$. C is called the square root of Λ^{-1}. Subsequently C is

$$C = \begin{bmatrix} \sqrt{w_1} & & & \\ & \sqrt{w_2} & & 0 \\ & & \ddots & \\ 0 & & & \sqrt{w_n} \end{bmatrix} \tag{5.11}$$

which can be transformed into an ordinary least squares problem by multiplying both sides of Eq. 5.8 by C, which leads to

$$C\theta_L = C\Phi\theta_R + C\varepsilon_R \tag{5.12}$$

one can then define $Z = C\theta_L$, $W = C\Phi$ and $\delta = C\varepsilon_R$, leading to

$$Z = W\theta_R + \delta \tag{5.13}$$

with the ordinary least squares solution being

$$\hat{\theta}_R = \left(W^T W\right)^{-1} W^T Z \tag{5.14}$$

An ordinary or unweighted least squares problem is therefore obtained from a weighted least squares problem through transformation of all the involved variables. Each observation of dependent and independent variables, and the intercept term (which is a column of ones, added to the matrix of independent variables) has to be multiplied by the square root of the weight for that point $\sqrt{w_i}$. Estimates of the parameters, tests, confidence intervals, and residuals can then be derived using ordinary least squares as described earlier (Weisberg, 1980, p. 76; Helsel and Hirsch, 1992, p. 281).

Thus, the uncertainty introduced by the varying identifiability of the local model parameters can be considered using this approach.

5.3.3.6 Conclusions

A range of regional model structures is presented above. Univariate regression is most commonly used and is appealing due to its simplicity. It also allows for estimates of uncertainty in its parameters and predictions. Additionally, the use of weighted regression seems to be a promising way to include local parameter uncertainty.

A seemingly unrelated regression approach allows for the consideration of correlation between the local parameters. This is a desirable char-

acteristic, and applications of these approaches to simple two parameter models can be found in the literature. However, the extension of this approach to more complex model structures is beyond the scope of the research presented in this study.

Emerging technologies such as generalized and symbolic regression offer interesting alternatives to relax the *a priori* assumption of a particular functional relationship between local model parameters and catchment characteristics. However, the use of symbolic regression means making the second step before the first; other more important issues such as the problem of local parameter estimation should be addressed first. The use of a GRNN might be a way to test whether there are underlying non-linear relationships.

5.3.4 Regional procedure

Some variations on the conventional regionalization procedure have appeared recently in the literature (Funke *et al.*, 1999; Vogel *et al.*, 1999; Lamb *et al.*, 2000; Yu and Yang, 2000).

The traditional regionalization procedure, as outlined in Sec. 5.2, separates the calibration of the individual local model structures from the calibration of the regional model structure. Local parameter sets are selected in such a way that they represent individual catchments as well as possible, usually based on a single OF. However, as mentioned above, different parameter sets are often equally acceptable representations of a single catchment, especially when a single-objective optimization procedure is used. This is one reason that different researchers have used variations on the traditional regionalization procedure. Another is the recognition that optimal local model performance may not necessarily give rise to an optimal regional model. It might therefore be advantageous to consider both local and regional performance simultaneously.

5.3.4.1 Combined procedure

One variation lies in the definition of the overall calibration objective for the individual local model structures. Vogel *et al.* (1999) and Fernandez *et al.* (2000) derived an overall OF which combines the performance as-

sessment of each local model, based on the abcd model structure (see for example Alley, 1984), with the performance of the regional model structure in relating catchment characteristics and local model parameters. "Instead of choosing parameters which minimize the model residuals alone, our goal is to both minimize the model residuals and maximize the goodness-of-fit relations between model parameters and basin characteristics, concurrently" (Vogel *et al.*, 1999). The overall OF to be maximized can be written as follows,

$$\text{maximize } \frac{1}{n}\sum_{i=1}^{n} R_i^2 + \frac{1}{p}\sum_{j=1}^{p} R_j^2 \tag{5.15}$$

where n is the number of catchments and R_i^2 the coefficient of determination (see Eq. 5.26) for each local model, while p is the number of parameters and R_j^2 is the coefficient of determination associated with the regression model explaining each parameter individually. This approach attempts to select those local parameter combinations that have a good fit with respect to the regional model. The result of this study is that extremely good regional relationships can be derived with this method, while local model performances in the gauged catchments are generally only slightly inferior to the traditionally derived ones. However, the performance of the local models in the ungauged catchments, with parameters derived from the traditional and the improved regional models, were similar; no improvement in the regionalized model was found.

A somewhat similar approach was suggested by Funke *et al.* (1999) for a distributed model structure that contains six conceptual parameters. Parameter sets were found using a global optimization algorithm based on an NSE OF (Table 3.1). The optimum parameter sets with respect to streamflow showed no significant correlation with the characteristics of seven catchments. Sensitivity analysis revealed that considerable variation in the parameters is possible with only marginal loss of performance. The authors subsequently related two conceptual parameters to physical catchment characteristics based on physical reasoning, leaving four parameters to be estimated. The remaining parameters then showed a higher correlation to catchment characteristics when the optimization was repeated. However, in an additional iteration, a new parameter set

was only accepted if the identified basic regional relationship for the parameters was not violated. This improved the regression fit further.

The problem with these two approaches can be that, instead of finding the most appropriate value for a particular parameter in a catchment, one uses its insensitivity with respect to a certain OF to derive a regional relationship which might be logical, but is not actively supported by the data.

5.3.4.2 Sequential procedure

Another approach to overcome the problem of lack of identifiability of local model parameters and to relax the assumption of independence of the local parameters in the regional models is a sequential regionalization procedure suggested by Lamb *et al.* (2000; Lamb and Calver, 2002). This procedure starts in the traditional way by calibrating the local model structure to each of the gauged catchments. The parameter with the largest impact on model performance (averaged over all catchments) is then selected and a regional model derived. This parameter is subsequently set at the value derived from the regional model for each of the gauged catchments, while the local calibration (URS) for all catchments is repeated and the regional model for the next parameter is derived. This way, the remaining parameters tend to become more identifiable and the problem of selecting the optimum local value for the regional analysis is reduced. The procedure results in a much-improved performance of the regionalized model in terms of flood frequency estimation compared to conventional univariate regression.

5.3.4.3 Updating procedure

Another possibility is available in cases where a gauging station has recently been established in a catchment and only a short measurement series is available. Thiemann *et al.* (2001) introduced a Bayesian updating approach that conditions a parameter distribution as new data progressively become available. Their methodology, termed Bayesian Recursive Estimation (BARE), starts with an initial guess of the region where good parameter values are located. New data are then used as they

become available and a one-step-ahead prediction is made. One problem with this approach is that the parameter sets are (at least initially) dependent on the climatic regime of the period for which data is available. This can only be avoided if longer time-series are used (e.g., Yapo *et al.*, 1996). The case study provided by Thiemann *et al.* (2001) showed that the estimates for each parameter ultimately collapsed onto a single point. Such a point estimate suggests that no uncertainty in the estimated parameter value is left, a fact that might or might not be appropriate for the modelling purpose at hand. Research on adjustments to avoid this effect is currently underway. See the discussion between Beven and Young (2003) and Gupta *et al.* (2003) on this issue.

5.3.4.4 Conclusions

The combined approaches suggested by some authors are difficult to justify in the absence of a theory to support the derived relationships. It seems more appropriate to improve the identifiability of the local parameters using MOs, for example.

The approach by Lamb *et al.* (2000) seems to do this. The dimensionality of the calibration problem is stepwise reduced, which should make the remaining parameters more identifiable. This assumption and the usefulness of the sequential procedure are tested in the application example. The use of an updating technique for a recently gauged catchment is beyond the scope of this monograph and will not be investigated here.

5.3.5 Flow prediction – including uncertainty

The remaining step is to use regional models to estimate parameter values for an ungauged catchment and then run the local model with the available meteorological forcing data (precipitation and temperature or PE) to produce an estimate of the runoff hydrograph.

Chapter 4 provided a detailed account of the uncertainties involved in local modelling. These are of course propagated into the results of any regionalization and the uncertainties originating from the regional approach itself have to be added. Despite this, only a few studies have con-

sidered the implications of uncertainty in the regionalization procedure; examples are Yeh *et al.* (1997) and (to some extent) Seibert (1999b).

Yeh *et al.* (1997) regionalized the two parameters, N and K, of Nash's Instantaneous Unit Hydrograph (IUH) model for 42 catchments in Taiwan (Tung *et al.*, 1997). They considered the uncertainty related to the two regionalized parameters and their influence on the predicted IUH. The methods of uncertainty analysis tried by these authors are the point estimation methods of Rosenblueth (1981), and Harr (1989). Seibert (1999b) found that acceptable values of some of the parameters for the HBV model structure are widely distributed over their feasible range with respect to the Nash-Sutcliffe efficiency measure. He therefore replaced the best parameter values with the median values in each group, which reduced the correlation between catchment characteristics and some of the parameters. Seibert (1999b) also used a fuzzy criterion to combine different OFs to reduce the range of feasible values. Both studies concluded that the uncertainty in any regional procedure is too large to be ignored.

5.3.5.1 Monte Carlo

One approach in considering uncertainty is to use classical regression estimates of the uncertainty in the regression (regional) parameters. Standard analytical expressions are available if the assumption that the regression residuals are normally distributed with constant variance holds. Uncertainty in local parameters could additionally be included using the weighted regression approach described above. A Monte Carlo approach can then be used to sample from the (normal) regional parameter distributions and produce (normal) local parameter distributions for the ungauged catchment (assuming there are independent local parameters and that normal distributions are suitable to provide initial estimates). Local parameter values that are outside the feasible range can be assigned an *a priori* support of zero. This Monte Carlo approach produces a flow range for the ungauged catchment instead of a single (point) prediction for every flow value.

The value predicted by the regression equation is the distribution mean. Following Kottegoda and Rosso (1997, p. 362), the sum of squared errors SS_E can be derived in matrix form,

$$SS_E = \theta_L{}^T \theta_L - \hat{\theta}_R \Phi^T \theta_L \tag{5.16}$$

The error variance is the SS_E divided by the degrees of freedom, *i.e.,* the number of gauged catchments (calibrated parameter values) used n minus the number of regression parameters, p, to be estimated,

$$\hat{\sigma}^2 = \frac{\theta_L{}^T \theta_L - \hat{\theta}_R \Phi^T \theta_L}{n - p} \tag{5.17}$$

Additionally, the $100(1-\alpha_{CFL})$ confidence limits on the (estimated) mean response $\mu_{\hat{\theta}_L|\phi}$ can be calculated as follows (Kottegoda and Rosso, 1997, p.381–383):

$$\mu_{\hat{\theta}_L|\phi} \pm t_{n-p,\alpha_{CFL}/2} \sqrt{\hat{\sigma}^2 \phi \left(\Phi^T \Phi\right)^{-1} \phi^T} \tag{5.18}$$

where Φ is a matrix of catchment properties of the gauged catchments used to derive the regional model, ϕ is a vector of catchment characteristics for the new (ungauged) catchment, and t is the ordinate of the student t-distribution.

Lamb (Centre of Ecology and Hydrology, Wallingford, UK, personal communication) and colleagues compared a standard regression uncertainty approach (based on normal distributions) to a new bootstrap method. They had very mixed results. However, the new approach potentially allows more information about the local parameters to be included in the analysis and is more general with respect to distributional assumptions.

5.3.5.2 Fuzzy regression

Fuzzy Regression (FR) is an alternative approach to standard statistical approaches. It can be either linear (Savic and Pedrycz, 1991) or non-linear (Bardossy *et al.*, 1990). Özelkan and Duckstein (2000) give a list of cases where FR might be superior to statistical regression: (1) when statistical regression analysis is not supported due to small sample size, this is often the case in regionalization studies where the number of

catchments is often limited (e.g., Magette *et al.*, 1976 (16 catchments); Servat and Dezetter, 1993 (20); Seibert, 1999b (11), Funke *et al.*, 1999 (7) etc.); (2) when assumptions about statistical distributions (usually normal) cannot be justified; (3) when the aptness of the regression model is poor; or (4) if human judgement is involved.

It is particularly useful when a small number of (possibly inaccurate) data points precludes classical regression. There are, however, limitations due to its potential sensitivity to data outliers (Özelkan and Duckstein, 2000). Other problems are partly related to this; for example, not all data points are always considered when the FR parameters are estimated and prediction intervals become wider if more data are collected. However, regionalization studies are particularly likely to contain outliers as will be shown in the application example in this chapter. Özelkan and Duckstein (2000) describe a multi-objective approach to FR to reduce some of those problems.

The advantage of FR being suitable even for small sample sizes seems appealing in the context of this research. However, a much larger sample (more catchments) has to ultimately be available for a sensible analysis. Considering also the high likelihood that outliers will be present in the analysis, the standard regression approach seems more promising and fuzzy regression is not investigated further.

5.3.5.3 Conclusions

The Monte Carlo approach using standard estimates of uncertainty around the regional parameter estimates seems to be a suitable first step in introducing estimates of uncertainty into regional flow predictions. It will therefore be included in the regional framework suggested in Sec. 5.4.

5.3.6 Uniqueness of place

There can be little debate about the fact that flow processes in natural catchments are very complex, and that the heterogeneity of soil and vegetation and the presence of preferential flowpaths, e.g., macropores,

create a particular uniqueness for every location. Beven (2000b) discusses this uniqueness and its effect on modelling practice or philosophy.

The problem of outliers, *i.e.,* unexpected model/catchment behaviour, is very important in the context of regionalization studies. The task is to discover the reasons why catchments behave differently. Why does a certain model structure (Beven, 2000b) after calibration underestimate flow in one catchment, and overestimate it in another? The uniqueness of catchments are revealed as model residuals in such cases. This problem will be demonstrated in the application example in Sec. 5.5.

Seibert (1999b), for example, found a high negative linear correlation between percentage forest and one of the recession coefficients of the HBV model structure. However, the second largest coefficient and the smallest one were derived for catchments of almost identical forest areas (87% to 84.5%). This example illustrates the importance of analysing the reliability and therefore the uncertainty related with the regression equations derived. Another (unaccounted for) catchment attribute seems to influence the latter catchment, while having little effect on the remaining ones.

Another example is the study by Funke *et al.* (1999). They adjust outliers, in this case parameter values found for one or two out of seven catchments that do not fit the general *correlation trend*, to derive better regression equations. The adjustment is done by recalibration of the outliers using reduced feasible parameter ranges. The new ranges are derived from the results in the remaining catchments. This results in much improved regional relationships with only limited loss of performance in individual catchments. However, the result is basically an *a priori* assumed relationship and any informative value initially attached to the outliers is lost.

Weisberg (1980, p.114) suggests the following in case an outlier is perfectly legitimate, e.g., not caused by measurement error or unsuccessful model calibration and nothing exceptional has occurred: "However, the model for Y for this combination of X-values does not conform to the line or plane that describes (most of) the cases. This case may then be the most important in a study, as it could represent new and unexpected information. The researcher may wish to study the condition of this point separately. However, a linear model may still be appropriate for the rest

of the data so that the analysis might often proceed by deleting this case." It seems therefore sensible to accept outliers and analyse them closely in an attempt to explain them, but possibly exclude them from the derivation of a regional model. This requires accepting that the regional model might sometimes fail to give a sensible local parameter estimate for a particular catchment. It is a reasonable approach as long as it is unknown why a particular (outlier) catchment behaves differently.

The example by Funke *et al.* (1999) is closely related to the aspect of catchment uniqueness which is also reflected in the model parameters. Every catchment requires a different parameter set to simulate its response. These parameters are effective values, integrating the influence of preferential flow paths, for example, which is not easily measured in the field. There is also the problem of parameter ambiguity, or model equifinality (e.g., Beven, 2000b): it is often not possible to discriminate between competing parameter sets for a particular case.

The uniqueness of catchments has been recognized in this section. Outliers in regionalization studies are important indicators of this uniqueness; their role in deriving regional models should therefore be analysed carefully.

5.4 A Framework for Regionalization

The regional framework proposed here can be divided into four stages: catchment selection and analysis, local model selection and application, regional model development and application, and prediction (Fig. 5.3).

The available catchments should be clustered in such a way that an individual cluster is represented by a single model structure. This clustering could, for example, be based on variables that describe the subsurface features of the catchment if geology is a dominant feature for those catchments selected. This variable could be the baseflow index. In a UK setting, this has the additional advantage that an estimate of the baseflow index for any ungauged catchment can be derived using the HOST system – which is then called BFIHOST – as described earlier (Boorman *et al.*, 1995). Different local model structures may have to be considered for different clusters. It has to then be ascertained whether the

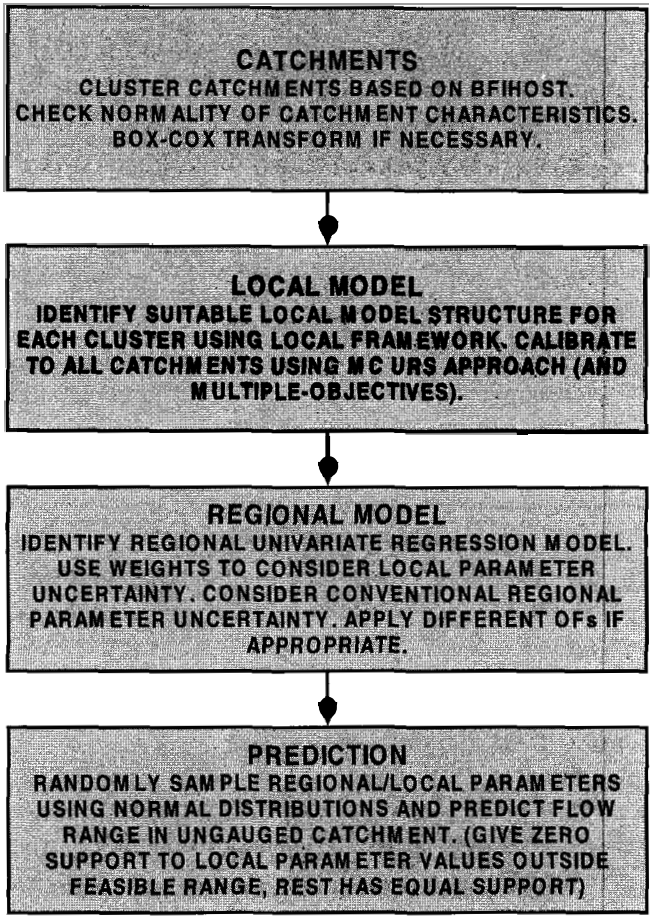


Fig. 5.3 Regional modelling framework.

catchment characteristics in each cluster are normally distributed (mainly for the consideration of uncertainty at a later stage). If not, then a Box-Cox transformation (Chapter 3) can be applied. The parameter λ of this transformation can either be guessed (the usual approach) or, in a more objective approach, optimized using the fit of the variables to a straight line in a normality plot.

At least one suitable local model structure(s) has to be identified for each cluster in the next stage. This can be done using the local frame-

work of corroboration and rejection introduced in Chapter 4. The selected model structure(s) can then be *calibrated* to each catchment using a Monte Carlo uniform random sampling approach. Again, the assumption of normal distribution should be tested.

The next stage is the identification of a regional univariate regression model. This includes calculating the empirical identifiability measure (introduced in Chapter 4) for each local model parameter as an average value over all catchments and using it as a regression weight to consider local parameter uncertainty. One can transform the dependent (local parameters) and/or independent variables (catchment characteristics) to consider differences in magnitude and non-linearity. This includes a conventional estimate of regression parameter uncertainty. The sequential approach (suggested by Lamb and Calver, 2000) applied in an objective way using the identifiability measure is an option at this stage. Three to five parameters can be identified using a single OF. There is therefore no need to regionalize only one parameter in every iteration.

The last stage uses random samples of regression parameters from normal distributions to calculate local parameter values, while giving a support of zero to local parameter values which lie outside of their feasible range. Finally, random samples, based on normal distributions, are drawn from those distributions to produce local parameter sets and apply them to estimate a flow range in the ungauged catchment.

This framework, and some variations on it, will be examined in the following application example in order to test its suitability.

5.5 Application Example

The purpose of this relatively small-scale application example is to test some of the optional techniques discussed above and address some of the issues raised. Successful techniques can then be included in the framework suggested in Section 5.4 (Fig. 5.3). Additionally, some aspects, such as parameter identifiability, can better be investigated in a pilot-scale example.

The disadvantage of a small sample size, as shown here, is that the resulting regional models might not be very robust. The aim of this study

is therefore rather to test methodologies and various aspects that are important (but largely untested) for this type of statistical regionalization study. The final regional models should only be seen as indicators.

The structure of the example mirrors the structure of the preceding review section.

5.5.1 Data

Data for ten catchments located in the southeast of England are available (Table 5.1). The catchments are predominantly rural and of varying geologies, consisting mainly of chalk, clay and greensand. They range from impervious clay catchments, such as the Blackwater@Ower or the Eden@Penshurst, over more evenly mixed catchments, e.g., Medway@Teston, the Teise@Stonebridge or the Upper Medway@Chafford Weir, to the Test@Broadlands, a highly permeable chalk catchment.

Time-series of (naturalized) flow, precipitation, temperature and potential evapotranspiration (PE) ranging over different periods (Fig. 5.4) were utilized. All periods lie within the range of 01.01.1989 to 31.12.1996.

Flow duration curves show cumulative frequency of streamflow versus percentage of time this streamflow is exceeded, usually for daily flows. Their shape is an indicator of the catchments' response to rainfall input. The steeper the flow duration curve, the more variable the response (Dingman, 1994, p.14). "An initially steeply sloped curve results from a very variable discharge, usually from small catchments with little storage where the stream flow reflects directly the rainfall pattern. Flow duration curves that have a very flat slope indicate little variation in flow regime, the resultant of the damping effects of large storage." (Shaw, 1994, p.295)

Responses of the different catchments can be compared by plotting their flow duration curves into a single plot after dividing them by the mean discharge of the river (Linsley *et al.*, 1949). The result is shown in Fig. 5.5. The plot shows the flat curves of the baseflow-dominated rivers such as the Test@Broadlands and the Medina@ShideWeir, while faster responding catchments like Eden@Penshurst, Medway@Teston or Eastern Rother@Udiam have much steeper curves. This response character-

Table 5.1 General catchment characteristics.

River	Location	AREA[1] [km^2]	Q_{mean}[2] [m^3/sec]	Q_{95}[2] [m^3/sec]	Q_{10}[2] [m^3/sec]	61-90 Av. An. Rain-fall[2] [mm]
Blackwater	Ower	102.4	0.85	0.154	2.131	836
Eden	Penshurst	224.8	1.76	0.232	3.853	742
East. Rother	Udiam	204.7	2.08	0.172	5.312	857
Medway	Teston	1261.3	10.56	1.526	23.840	744
Teise	Stone Bridge	134.5	1.31	0.202	2.553	810
Up. Medway	Chafford Weir	252.1	2.99	0.533	6.395	830
Eastern Yar	Burnthouse	59.6	0.41	0.040	0.846	
Medina	Shide Weir[3]	28.5	0.28	0.082	0.502	839
West. Rother	Hardham	360.7	4.96	1.694	9.829	899
Test	Broadlands	1035.9	10.93	5.780	16.560	790

[1] Data from FEH, 1999, CD.

[2] Data from http://www.nwl.ac.uk/ih/nrfa.

[3] This station was a flume structure at SZ 503 874. It was superseded in 1997 by an ultra-sonic gauging station at Upper Shide (SZ 5036 8810). The data used in this study were measured at the old station.

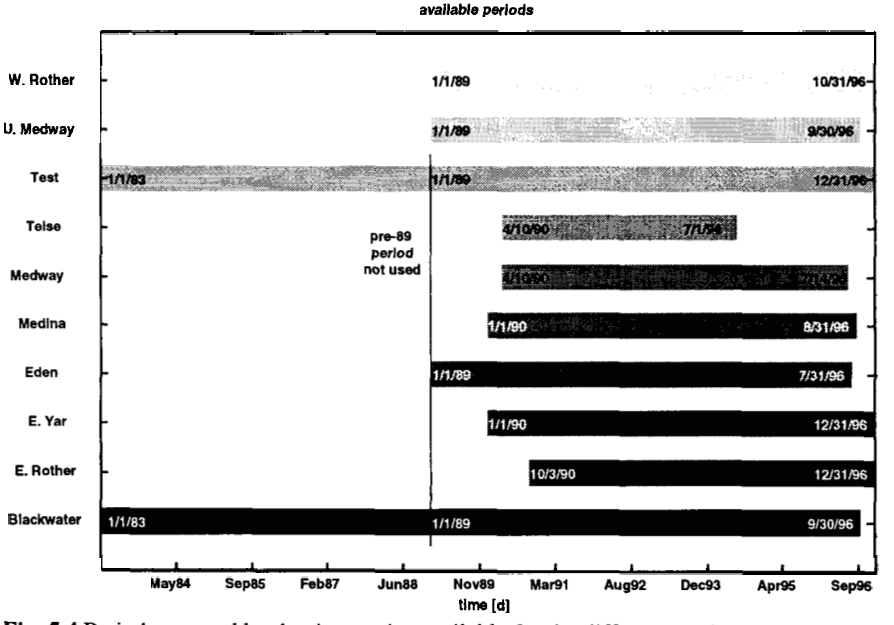


Fig. 5.4 Periods covered by the time-series available for the different catchments.

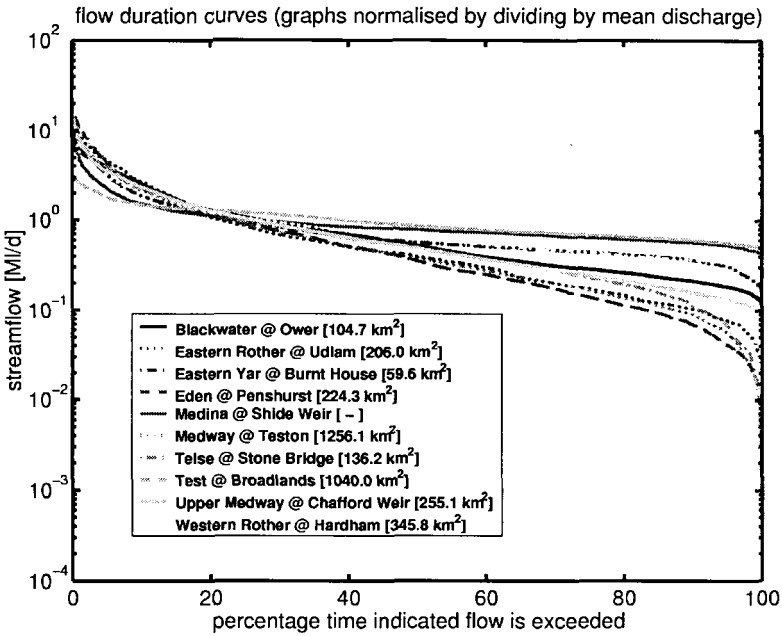


Fig. 5.5 Comparison of flow duration curves for the south-east England data set.

istic is also reflected in the baseflow index (BFIHOST, Tables 5.2 and 5.6).

Apart from being defined by the storage characteristics of the catchment, the shape of the flow duration curve is also influenced by topography, vegetal cover, land use and precipitation (Linsley *et al.*, 1949, p.585).

5.5.2 Catchment characteristics

All catchment characteristics are taken from the data CD that accompanies the Flood Estimation Handbook (FEH). These characteristics are suggested to be particularly suitable to describe a catchment's hydrological behaviour. The FEH catchment characteristics belong to the following four groups:

- Landform descriptors (AREA, LDP, DPLBAR, DPSBAR, ALTBAR, ASPBAR, ASPVAR)
- Index for attenuation effects attributable to reservoirs and lakes (FARL)
- Climate and soil descriptors (SAAR, RMED, SPRHOST, BFIHOST, SMDBAR, PROPWET)
- Urban and suburban land-cover descriptors ($URBEXT_{1990}$, URBLOC, URBCONC)

Table 5.2 explains the abbreviations used for the catchment characteristics. The variables for the catchments considered here are listed in Table 5.3 and statistical properties are summarized in Table 5.4.

Some of these catchment characteristics are highly correlated, e.g., Baseflow index (BFIHOST) and standard percentage runoff (SPRHOST) are both measures of a catchment's runoff response to rainfall input. It is therefore necessary to analyse this correlation and reduce the number of

Table 5.2 Description of catchment characteristics (Bayliss, 1999).

Characteristic	Unit	Description
AREA	km^2	Catchment drainage area
LDP	km	Longest drainage path
BFIHOST	-	Baseflow index derived using the HOST classification
SPRHOST	%	Standard percentage runoff derived using the HOST classification
FARL	-	Index of flood attenuation due to reservoirs and lakes
PROPWET	-	Index of proportion of time that soils are wet
DPLBAR	km	Index describing catchment size and drainage path configuration
DPSBAR	mkm^{-1}	Index of catchment steepness
ASPBAR	-	Index representing the dominant aspect of catchment slopes
ASPVAR	-	Index describing the invariability in aspect of catchment slopes
RMED-1D	mm	Median annual maximum 1-day rainfall
RMED-2D	mm	Median annual maximum 2-day rainfall
RMED-1H	mm	Median annual maximum 1-hour rainfall
SAAR	mm	1961-90 standard-period average annual rainfall
$SAAR_{4170}$	mm	1941-70 standard-period average annual rainfall
$URBEXT_{1990}$	-	FEH index of fractional urban extent for 1990
URBCONC	-	Index of concentration of urban and suburban land cover
URBLOC	-	Index of location of urban and suburban land cover

Table 5.3 Detailed catchment characteristics (Source: FEH CD).

River @ Location	Blackwater @ Ower	Eden @ Penshurst	E Rother @ Udiam	Medway @ Teston	Teise @ Stone Bridge
AREA	102.38	224.82	204.66	1261.33	134.46
LDP	18.48	33.92	30.93	65.99	23.68
BFIHOST	0.479	0.425	0.388	0.439	0.443
SPRHOST	34.2	41.2	44.4	41.3	42.6
FARL	0.985	0.925	0.975	0.949	0.905
PROPWET	0.33	0.35	0.35	0.35	0.36
DPLBAR	11.06	20.33	17	36.03	13.45
DPSBAR	44.4	47.2	92.6	53.6	78.3
ASPBAR	124	136	112	81	57
ASPVAR	0.17	0.11	0.11	0.05	0.1
RMED-1D	35.7	32.9	36.6	33.3	35
RMED-2D	46.3	44.1	48.8	44.3	47.9
RMED-1H	10.7	11.4	11.6	11.7	11.8
SAAR	837	742	857	744	812
$SAAR_{4170}$	867	764	861	755	809
$URBEXT_{1990}$	0.009	0.016	0.008	0.019	0.005
URBCONC[1]	0.327	0.556	0.508	0.614	-
URBLOC[1]	0.875	1.217	0.976	0.965	-

[1]These variables are not considered further, because their values for Teise @ Stonebridge are not available.

Table 5.3 (cont.)

River @ Location	Up Med@ Chafford Weir	East. Yar @ Burnthouse	Medina @ Shide Weir	W Rother @ Hardham	Test @ Broadlands
AREA	252.05	59.58	28.54	360.71	1035.93
LDP	29.09	19.49	11.41	56.19	69.89
BFIHOST	0.441	0.743	0.753	0.666	0.898
SPRHOST	42.4	24.3	23.6	27.4	9.4
FARL	0.938	0.992	0.985	0.973	0.964
PROPWET	0.35	0.33	0.33	0.34	0.34
DPLBAR	14.36	10.68	5.84	30.13	39.97
DPSBAR	82.4	84.6	78.5	72.9	50.1
ASPBAR	28	6	59	120	176
ASPVAR	0.06	0.06	0.1	0.1	0.15
RMED-1D	34.8	33.9	35	39.4	33.3
RMED-2D	47.3	45.3	46	50.8	43.1
RMED-1H	11.7	9.4	9.5	10.4	10.6
SAAR	831	844	839	899	790
$SAAR_{4170}$	852	910	911	918	818
$URBEXT_{1990}$	0.02	0.017	0.015	0.008	0.01
URBCONC	0.601	0.44	0.342	0.497	0.56
URBLOC	1.231	0.895	0.847	1.16	0.841

characteristics considered. An alternative approach is the derivation of principal components (Gershenfeld, 1999, p.136–138). Principal component analysis is the transformation of an existing data set into new uncorrelated variables using the covariance matrix of the original data. Sefton and Howarth (1998) for example use this approach to derive new variables such as topography and soils/geology from the initial set of catchment characteristics. However, the newly derived variables reduce the ease with which regional relationships can be interpreted. This approach is therefore not considered here. However, the correlation analysis and subsequent reduction of variables has to be treated with caution due to the small number of catchments available.

Highly (linearly) correlated catchment characteristics (*i.e.*, with correlation coefficient above 0.7) have been highlighted in Table 5.5. They are also listed below and a representative variable for each group of correlated characteristics is selected. Underlined variables have correlation coefficients equal to or above 0.9. Brackets indicate a negative coefficient:

- AREA LDP DPLBAR
- BFIHOST (SPRHOST) (RMED-1H)
- FARL (PROPWET) (RMED-1H) $SAAR_{4170}$
- ASPBAR ASPVAR
- RMED-1D RMED-2D SAAR
- RMED-1H (BFIHOST) SPRHOST (FARL)
  PROPWET ($SAAR_{4170}$)
- SAAR RMED-1D RMED-2D $SAAR_{4170}$

Consequently:

- DPLBAR is an index that combines catchment size and drainage path configuration. AREA and LDP can therefore be dropped.
- BFIHOST and SPRHOST are highly correlated, therefore only BFIHOST is considered for regional modelling. However, both are kept for clustering purposes.
- A strong correlation between RMED-1D and RMED-2D means that only the latter is considered further.
- ASPBAR and ASPVAR describe different physical aspects of the catchment and are therefore both retained even if they are statistically correlated.

Table 5.4 Univariate statistical properties of catchment characteristics.

Characteristic	Mean	Median	Standard deviation	Minimum	Maximum
AREA	366.5	214.7	426.9	28.5	1261.3
LDP	35.9	30.0	20.7	11.4	69.89
BFIHOST	0.568	0.461	0.180	0.388	0.898
SPRHOST	33.08	37.7	11.5487	9.40	44.40
FARL	0.959	0.969	0.029	0.905	0.992
PROPWET	0.343	0.345	0.011	0.33	0.36
DPLBAR	19.89	15.68	11.59	5.84	39.97
DPSBAR	68.46	75.60	17.78	44.40	92.60
ASPBAR	89.90	96.50	52.79	6.00	176.00
ASPVAR	0.101	0.100	0.038	0.05	0.17
RMED-1D	34.99	34.90	1.94	32.90	39.40
RMED-2D	46.39	46.15	2.36	43.10	50.80
RMED-1H	10.88	11.05	0.91	9.40	11.80
SAAR	819.5	834.0	49.19	742	899
$SAAR_{4170}$	846.5	856.5	59.03	755	918
$URBEXT_{1990}$	0.0127	0.0125	0.0053	0.005	0.02

Table 5.5 Correlation matrix for catchment characteristics.

	AREA	LPD	BFI-HOST	SPR-HOST	FARL	PROP-WET	DPLBAR	DPSBAR
AREA	1.00							
LPD	0.91	1.00						
BFIHOST	0.15	0.21	1.00					
SPRHOST	-0.21	-0.26	-0.68	1.00				
FARL	-0.13	-0.13	0.51	-0.53	1.00			
PROPWET	0.23	0.25	-0.64	0.66	-0.88	1.00		
DPLBAR	0.90	1.00	0.22	-0.28	-0.14	0.25	1.00	
DPSBAR	-0.47	-0.41	-0.04	0.19	0.12	0.12	-0.46	1.00
ASPBAR	0.40	0.55	0.12	-0.27	0.01	0.02	0.06	-0.64
ASPVAR	-0.11	-0.02	0.19	-0.34	0.19	-0.28	0.04	-0.49
RMED-1D	-0.33	-0.04	-0.04	0.05	0.30	-0.11	-0.08	0.39
RMED-2D	-0.46	-0.20	-0.27	0.33	0.05	0.16	-0.25	0.60
RMED-1H	0.29	0.27	-0.79	0.75	-0.76	0.90	0.26	-0.11
SAAR	-0.54	-0.31	0.22	-0.18	0.54	-0.38	-0.35	0.61
$SAAR_{4170}$	-0.58	-0.43	0.46	-0.40	0.72	-0.67	-0.46	0.55
$URBEXT_{1990}$	0.20	0.00	-0.03	0.09	0.03	-0.12	-0.03	-0.05

- SAAR and $SAAR_{4170}$ are highly correlated. Only SAAR is kept because all time-series used a start in 1989 or later.
- PROPWET and RMED-1H produce a correlation coefficient of 0.9. Only PROPWET will be kept, since RMED-1H also shows relatively high correlations with some other characteristics. PROPWET is also highly correlated with FARL, which is also consequently dropped.

The catchment characteristics subsequently used for regional analysis are therefore:

- Landform: DPLBAR, DPSBAR, ASPBAR, ASPVAR.
- Climate and soil: BFIHOST, PROPWET, SAAR, RMED-2D.
- Urban and suburban: $URBEXT_{1990}$.

A more detailed explanation of the selected variables follows.

Landform

Four landform descriptors are kept. They are derived from the Institute of Hydrology Digital Terrain Model (IHDTM), a regular nodal grid of 50m x 50m elements (Bayliss, 1999, p.1). These are DPLBAR – the mean of

Table 5.5 (cont.)

	ASPBAR	ASPVAR	RMED – 1D	RMED – 2D	RMED – 1H	SAAR	SAAR 4170	URBE XT$_{1990}$
AREA								
LPD								
BFIHOST								
SPRHOST								
FARL								
PROPWET								
DPLBAR								
DPSBAR								
ASPBAR	1.00							
ASPVAR	0.75	1.00						
RMED-1D	0.07	0.14	1.00					
RMED-2D	-0.17	-0.07	0.93	1.00				
RMED-1H	0.17	-0.09	-0.10	0.10	1.00			
SAAR	-0.20	0.10	0.65	0.80	-0.42	1.00		
$SAAR_{4170}$	-0.31	0.06	0.64	0.54	-0.73	0.91	1.00	
$URBEXT_{1990}$	-0.45	-0.65	-0.54	-0.48	-0.08	-0.42	-0.18	1.00

the distances between each node on a 50m grid and the catchment outlet in km (a measure of catchment size and configuration), DPSBAR – the mean of all inter-nodal slopes for the catchment in mkm^{-1} (a measure of overall steepness), ASPBAR – the mean direction of all inter-nodal slopes in the catchment in degrees with north being zero and values increasing clockwise up to 360° (a measure of the dominant aspect of-catchment slopes), and ASPVAR – a measure of the invariability of slope directions, of which a value close to zero indicates considerable variability, while a value close to one indicates that the slopes show a tendency to face a similar direction.

Climate and soil

Four climate and soil variables are considered: two are related to soil and two to rainfall characteristics. They are BFIHOST – the baseflow index derived using the HOST classification (Boorman *et al.*, 1995); PROPWET – the proportion of time that the (simulated) soil moisture deficit (SMD) was below or equal to 6mm during 1961–1990 (soils are at field capacity if SMD = 0, however the threshold value was introduced to account for the theoretical nature of the SMD value; Bayliss, 1999, p. 33–34); SAAR – the standard period (1961–1990) average annual rainfall in mm supplied by the Meteorological Office on a 1 km grid; and RMED-2D – the median annual maximum 2-day rainfall in mm, again for a 1 km grid (Bayliss, 1999, p.27).

Urban and suburban

Only one variable describing the level of urbanization in each catchment is used, $URBEXT_{1990}$ – the extent of urban and suburban land cover in 1990. However, all selected catchments available are essentially rural based on the categories of catchment urbanization given in Bayliss (1999, p.47). The maximum $URBEXT_{1990}$ value is only 0.02 as shown in Table 5.4.

5.5.3 Local model

The local model structures tested are a probability distribution of stores (pd4), a catchment wetness index (cwi) and a catchment moisture deficit

(cmd) model structure, all combined with a parallel routing component consisting of two linear stores (2pll). Details of the model components are given in Chapter 3. The initial conditions (model moisture states) are calibrated with some initial parameter estimates before the actual model parameters are estimated. The Teise@Stonebridge is excluded from the regional analysis to be used later as a test catchment.

Local modelling approach

A uniform random sampling procedure was used to derive 10000 parameter sets for each catchment. The OF applied is the RMSE measure (Table 3.1, p.81), unless stated otherwise.

Initial local model structure

The first structure that has been applied to all catchments is pd4-2pll. The structure consists of a probability distribution of soil moisture stores producing effective rainfall (ER) purely through an overflow mechanism. The actual evapotranspiration is a linear function of the average soil moisture content. The ER is divided by a constant split parameter into a quick and a slow component, and each component is routed through a linear reservoir. This structure contains five parameters: the maximum storage capacity cmax, the shape of the Pareto distribution of soil mois-

Table 5.6 RMSE calibration results for local model structure, pd4-2pll. Catchments are sorted according to BFIHOST values.

No.	Catchment	RMSE	BFIHOST
c01	EasternRother @ Udiam	0.91	0.388
c02	Eden @ Penshurst	0.70	0.425
c03	Medway @ Teston	0.57	0.439
c04	UpperMedway @ ChaffordWeir	0.84	0.441
c05	Teise @ Stonebridge	0.66	0.443
c06	Blackwater @ Ower	0.73	0.479
c07	WesternRother @ Hardham	0.52	0.666
c08	EasternYar @ Burnthouse	0.61	0.743
c09	Medina @ Shide Weir	0.79	0.753
c10	Test @ Broadlands	0.19	0.898

[1]Fit improves largely with unreasonable parameter values in this chalk catchment.

ture capacities b, the split parameter alpha (*i.e.,* percentage quick contribution), and the two residence times, k(quick) and k(slow).

The resulting RMSE (Table 3.1, p.81) values for this model structure being calibrated to all ten catchments are shown in Table 5.6. However, the best model (according to RMSE) for the Blackwater@Ower gives an unsatisfactory result in the way that it largely overpredicts the low flow periods in order to fit the high flows correctly (and thus yields a poor volumetric fit). There seems to be a problem with the data, resulting, for example, in unreasonable values for the parameter describing the split between quick and slow catchment responses. This catchment is therefore excluded from further analysis.

Alternative local model structures and models

Catchment c10, the Test@Broadlands, is a chalk catchment. The calibration result seems reasonable, with an NSE value of 0.78. However, a different model structure is more appropriate for this type of baseflow-dominated catchment, which also stands out as a separate cluster due to its high BFIHOST value. This structure is a combination of the Penman SMA module and the leaky routing component (see Chapter 3).

The leaky component differs from the one described in Chapter 3 in the sense that the lower threshold was set to zero. The original module was initially applied to the Lavant by Drane (2000). The Lavant is a chalk river located in south-central England. It is ephemeral, with flow ceasing for about five months a year and a very low runoff yield (12% of rainfall) (Drane, 2000). The lower threshold in the leaky routing module was necessary to simulate periods of zero flows, *i.e.,* all runoff contributes to the catchment losses. The Test@Broadlands showed no ephemeral behaviour and the threshold was therefore not required.

On the basis that it represents a different catchment type, catchment c10 was not included in any subsequent regional analysis. A further division into clusters is not undertaken here due to the small number of catchments available.

As mentioned earlier, research results (Kokkonen and Jakeman, 2001) suggest that identical components (and therefore parameters) used within different model structures can have different optimum parameters This is tested here by applying a metric model structure (cwi-2pll) and an

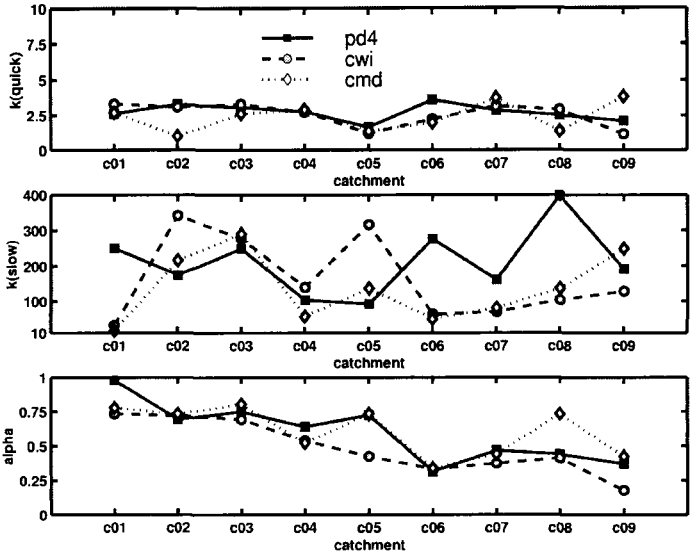


Fig. 5.6 Comparison of k(quick) [d], k(slow) [d] and alpha [-] values for the pd4-2pll, cwi-2pll and cmd-2pll model structures. Values shown are optimal with respect to the RMSE criterion.

additional parametric one (cmd-2pll), and comparing the resulting routing parameters.

The metric and the second parametric model structures are applied to all catchments using the same URS approach (10,000 samples) as before, and the best k(quick), k(slow) and alpha parameters for all structures are selected based on the RMSE criterion. The variation in optimum values is shown in Fig. 5.6.

One can see that with respect to the parameter k(quick), all model structures show a relatively high degree of similarity in optimum values. However, there seems to be a tendency for cwi to produce higher values. There are only two catchments, c06 and c09, where pd4 or cmd produce significantly higher values. Generally this parameter seems to show little dependency on the SMA module, however. This parameter is usually well identified.

The result for k(slow) is basically a confirmation that it is impossible to identify this parameter using the overall RMSE as OF. The optimum

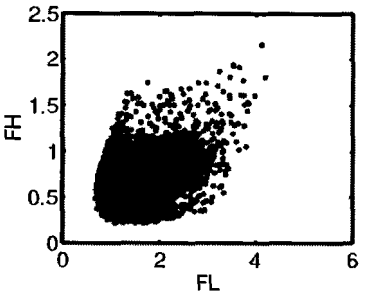


Fig. 5.7 Scatter plot showing a model population based on 10000 samples from uniform parameter distributions projected into a 2-dimensional OF space. The two OFs are the RMSE utilizing the residuals at time steps during which the observed flow is respectively above (FH) or below (FL) the mean flow.

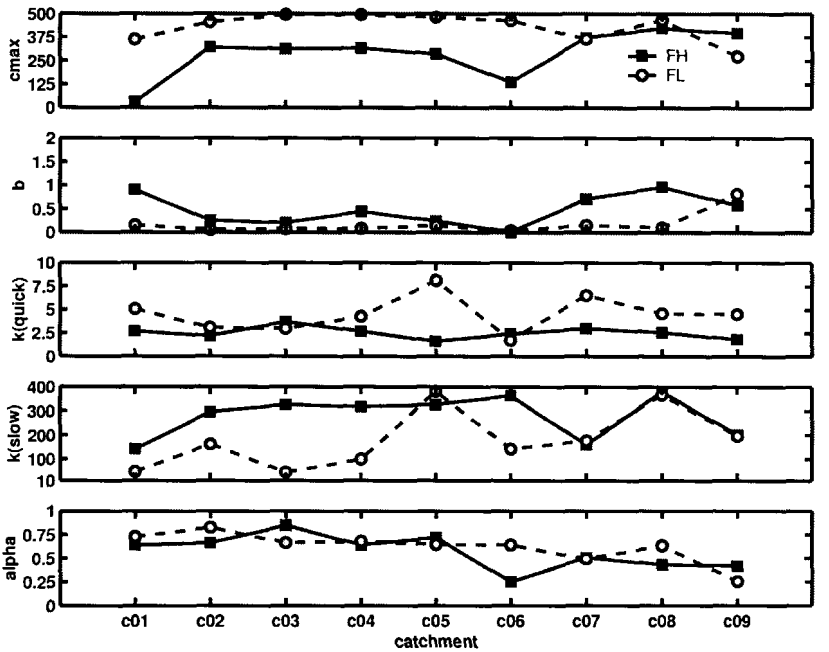


Fig. 5.8 Variation in optimum parameter values for two different OFs representing two different modelling purposes. The two OFs are the RMSE utilizing the residuals at time steps during which the observed flow is above (FH) or below (FL) the mean flow respectively. The model structure applied is pd4-2pll.

values vary widely and there is little structure (see also Fig. 5.11 and 5.12). The result is not sufficiently reliable to draw any conclusions.

The alpha values for the cwi module are consistently the lowest for all catchments, *i.e.*, a smaller alpha means a smaller contribution to quickflow and therefore a larger baseflow component. Alpha values for cwi also show an even more pronounced trend to decrease with increasing BFIHOST than the values for the other two SMA modules. This confirms the result of the study by Kokkonen and Jakeman (2001) with respect to this parameter. Figures 5.11 and 5.12 show that alpha is generally quite identifiable, which supports the validity of the result.

The probability distributed model structure (pd4-2pll) is selected for all subsequent analyses given its relatively good performance and parsimonious model structure; only two SMA parameters are required to be estimated. Before instigating regional modelling, we first investigate the effect of the choice of OF. Two very different OFs are applied to test the dependency of the local, and therefore possibly also the regional, model on the modelling purpose. Both OFs are based on the RMSE. The first, FH, assumes that the high flows are of particular importance, e.g., for flood frequency analysis (Lamb, 2000), while the second, FL, aims at optimising the reproduction of the catchment's low flow behaviour, e.g., for water resource purposes (Young, 2002). The OFs differ in that FH is the RMSE using only the residuals at time-steps when the observed flow is above a certain threshold, while FL only uses those below. The threshold chosen is the mean observed flow in this case. Figure 5.7 shows, for one example catchment, a pronounced trade-off between those criteria for the model structure applied here.

The best parameter sets (based on a URS with 10,000 points) for the two OFs for all catchments are shown in Fig. 5.8. One can see a considerable variation in values for almost all parameters in most of the catchments. It is therefore likely that the chosen modelling objective will have considerable influence on any regional analysis, *i.e.*, the derived regional model.

5.5.4 Regional model

Some of the possible regional model structures outlined earlier and selected in Section 5.4 are investigated here. The benchmark is a multiple univariate regression (here called *conventional regression*), which is the standard approach in regionalization studies as mentioned earlier.

The starting point is an analysis of the statistical characteristics of the variables involved and their correlation. This can already give some indication about what type of functional (regional) relationship might be appropriate. Caution is necessary in the case of this example, due to the small number of catchments available.

5.5.4.1 Variable properties and correlation analysis

The analysis of individual variables (calibrated model parameters and catchment characteristics) is a test of assumptions made during the regression analysis, mainly normality. The condition of normality is important if inference about the statistical properties of the estimated regression, e.g., confidence limits, is to be made at a later stage. If all variables in a regression are normally distributed, the resulting error distribution will likely also be. Normality plots are usually used to test this aspect.

Such a plot is shown, for all the catchment characteristics available, at least for the limited sample size represented here, in Fig. 5.9. One can see that all variables can be considered normal. The variable AREA might need a transformation due to two outliers, but is not used in any regional analysis in this section. Similar normality plots are used to analyse all parameter distributions in this example, but for reasons of space they are not shown here.

The correlation between variables is most easily tested by calculating a linear correlation coefficient, in the original and in log-transformed space (Table 5.7 and 5.8). A rank correlation coefficient can additionally be used to test for non-linear relationships (e.g., Tung *et al.*, 1997; Seibert, 1999b). Correlation coefficients and rank correlation coefficients cannot be compared directly in quantitative terms. However, cases where the rank correlation coefficient is significantly larger than the linear correlation coefficient can be indicative of a non-linear dependence. A non-

parametric test, *i.e.,* one that does not assume a particular functional form, is Spearman's rank correlation coefficient (e.g., Kottegoda and Rosso, 1998, p.280–282). In this test, two sets of variables, e.g., x_i and y_i (i = 1,2,...,n), are ranked separately such that the highest value of each variable is assigned a rank of 1 and the lowest a rank of n. The rank coefficient can then be estimated using

$$r_s = 1 - \frac{6\sum_{i=1}^{n} rd_i^2}{n\left(n^2 - 1\right)} \tag{5.22}$$

where rd_i is the difference in ranks between x_i and y_i.

In this context, it is also necessary to look at the correlation between the model parameters themselves. Sefton and Howarth (1998) for example found that the proportion of rainfall contributing to catchment storage and the time constant of catchment wetness decline were highly correlated (with a linear correlation coefficient of 0.92) and therefore used a linear regression relationship to derive one from the other. Tung *et al.* (1997) found a high correlation between the two Nash Instantaneous Unit Hydrograph (IUH) parameters N and K, which they later considered when deriving regional equations.

Table 5.7 Correlation coefficients for local pd4-2pll model parameters on normal and on log-transformed scale.

(a) Correlation coefficients for local parameters on original scale.

	cmax	b	k(quick)	k(slow)
b	-0.0080			
k(quick)	0.1939	-0.7817		
k(slow)	0.0400	-0.1770	-0.2279	
alpha	-0.6943	-0.5725	0.4557	-0.0594

(b) Correlation coefficients for local parameters on log-transformed scale.

	cmax	b	k(quick)	k(slow)
b	0.1560			
k(quick)	0.1688	-0.8216		
k(slow)	-0.1540	-0.1557	-0.1900	
alpha	-0.6211	-0.6492	0.5837	-0.0465

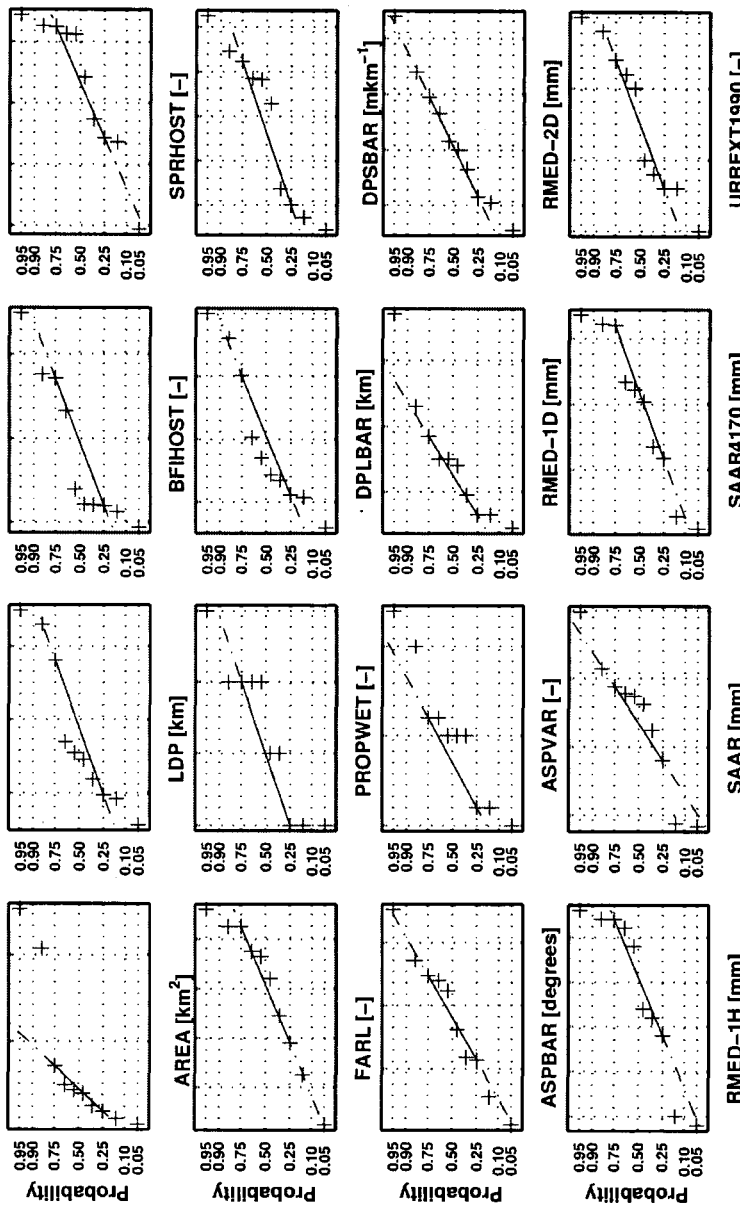


Fig. 5.9 Normality probability plots of available catchment characteristics. Only AREA appears to be possibly not normally distributed. A log-transformation can be used to change this. However, AREA is not used in the regression analysis later in the text.

In case of the pd4-2pll, a few correlation coefficients suggest some degree of correlation between the local parameters (Table 5.7). The largest is between b and k(quick) (-0.82 on log and -0.78 on normal scale). Also, alpha is related to b (and therefore k(quick)) and cmax with correlation coefficients between -0.57 and -0.69. The number of points on which these correlation coefficients are based is rather small and therefore these relationships are probably not very robust. However, one can see that this analysis should be part of any regionalisation procedure.

The correlation analysis between model parameters and catchment characteristics (Table 5.8) shows encouraging results for three parameters, but no correlations at the 5% level for cmax and for k(slow). These results have to be interpreted carefully due to the small sample size of eight catchments, since the Test@Broadlands could not be described by pd4-2pll and Teise@Stonebridge is kept apart for testing. Additionally, the problem of estimating k(slow) using the RMSE was already mentioned.

The only correlation above 0.5 found for parameter cmax is with $URBEXT_{1990}$ (Table 5.8). Again, this relationship might be of little importance since all catchments are essentially rural. A combination of several variables might therefore be required to explain this parameter in a regression model. The parameter could also be fixed to the median value of all calibration results as suggested by Seibert (1999b). A low correlation for this parameter could have been expected since no soil properties in the form of field capacity, porosity or wilting point were available. These are more likely to be related to cmax than the variables used here.

A similar lack of correlation for the water balance characteristics of the IHACRES model structure was reported by Sefton and Howarth (1998). Lamb *et al.* (2000) needed four catchment characteristics in a regression equation – topographic index, drainage path slope, $SAAR_{61\text{-}90}$, and urban area – to yield a coefficient of determination of 0.7 for cmax (using an hourly time-step for flood peak estimation).

A slightly better result is found for the shape parameter b. Three variables yield correlation coefficients above 0.5 on the original scale and one is even above 0.6 on log-transformed scale. These are the baseflow index BFIHOST (original: 0.56, log: 0.58), the index for catchment size and drainage path configuration DPLBAR (o.: -0.58, l.: -0.66), and an

index describing how often the soils are wet PROPWET (o.: -0.60, l.: -0.59). However, only the correlation to DPLBAR (on log-scale) is significant at the 5% level. The variable PROPWET is unlikely to be very useful in a regression since it only takes four different values (0.33, 0.34, 0.35 and 0.36) despite the large variation between the catchments.

Table 5.8 Correlation coefficients for all variables involved. Three catchments are excluded from this calculation: (1) Teise@Stonebridge, because it is kept as a test catchment; (2) Test@Broadlands, due to the differences in BFIHOST (see text); and (3) Blackwater@Ower, because no satisfactory calibration could be established judged by the visual fit. (The best parameter values are normally distributed, but show some outliers. Normality plot not shown.)

(a) Correlation coefficients for variables on original scale.

Characteristic	cmax	b	k(quick)	k(slow)	alpha
BFIHOST	0.4290	0.5595	-0.6585	0.3485	-0.9073**
DPLBAR	-0.0415	-0.5807	0.7228*	-0.1401	0.3370
DPSBAR	-0.3597	0.2267	-0.7376*	0.2130	-0.0291
PROPWET	-0.3446	-0.5945	0.7407*	-0.4618	[illegible]
SAAR	-0.1624	0.1765	-0.5874	0.0086	-0.3115
$URBEXT_{1990}$	0.5689	0.1508	0.1317	0.0159	-0.1781
ASPBAR	-0.4091	-0.3404	0.5389	-0.3785	0.4356
ASPVAR	-0.4319	0.0905	-0.0504	-0.2857	0.1255
RMED-2D	-0.3460	-0.0887	-0.2067	-0.3074	0.0214

*Significant at the 5% level.
**Significant at the 1% level.

(b) Correlation coefficients for variables on log-log-transformed scale.

Characteristic	cmax	b	k(quick)	k(slow)	alpha
BFIHOST	0.4688	0.5806	-0.6381	0.2815	-0.9554**
DPLBAR	-0.0337	-0.6615~	[illegible]	-0.0896	0.5475
DPSBAR	-0.3813	0.4472	-0.7098*	0.1016	-0.1751
PROPWET	-0.3376	-0.5934	0.7525*	-0.4227	[illegible]
SAAR	-0.2029	0.3363	-0.5600	-0.0425	-0.3792
$URBEXT_{1990}$	0.5912	0.2580	0.1137	-0.0961	-0.0971
ASPBAR	-0.4323	-0.3377	0.3830	-0.4016	0.4243
ASPVAR	-0.4074	-0.2182	-0.1181	-0.1657	0.0059
RMED-2D	-0.3655	0.0490	-0.1755	-0.3073	-0.0197

*Significant at the 5% level.
**Significant at the 1% level.
~This value is just outside the 5% significance level.

Table 5.8 (cont.)

(c) Spearman rank correlation coefficients for variables.

Characteristic	cmax	b	k(quick)	k(slow)	alpha
BFIHOST	0.2143	0.7857	-0.6071	-0.0357	-0.9643
DPLBAR	-0.0714	-0.6429		-0.1786	0.6429
DPSBAR	-0.2143	0.2857	-0.7500	0.4286	-0.0357
PROPWET	-0.3661	0.7768	0.5804	-0.4732	
SAAR	-0.2857	0.0357	-0.5000	0.1786	-0.2143
$URBEXT_{1990}$	0.4911	0.4196	0.1875	-0.1339	0.0089
ASPBAR	-0.2143		0.6786	-0.2857	0.4643
ASPVAR	-0.2054	-0.5446	0.0625	-0.0268	0.2054
RMED-2D	-0.4286	0.1071	-0.3571	-0.2500	-0.0714

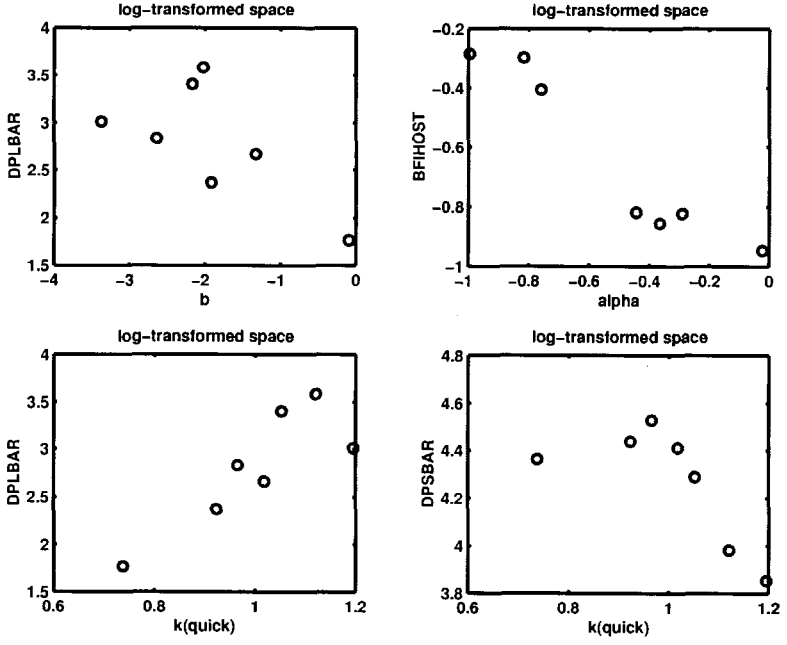


Fig. 5.10 Scatter plots for some of the correlations used for the conventional regression analysis plotted on log-log scale.

A value of zero for b is equal to a constant storage capacity over the catchment, while a value of 1 yields a uniform distribution of storage capacities. The results suggest higher b values for catchments with a larger contribution of baseflow, smaller catchment area and drier soils. However, all correlations are rather low, so a detailed investigation of this result is not useful at this stage.

The shape of the storage distribution function b is the only parameter for which the Spearman rank correlation coefficient gives considerably higher values than the linear correlation coefficients on normal or log-transformed scale (Table 5.8(c)). The highest rank correlation values are found with the variables ASPBAR (-0.82), BFIHOST (0.79) and PROPWET (-0.78). There is also a relatively high value with DPLBAR (-0.64). The fact that the (non-parametric) rank correlations are higher than those assuming a linear relationship is indicative of a possible non-linear relationship. The variable ASPBAR describes the mean aspect of the catchment. It is calculated as an average from the outflow direction (bearing) of each nodal point on the IHDTM within a catchment. It is therefore an indicator of the dominant aspect of catchment slopes. Its values increase clockwise from zero to 360°, starting from the north. A negative rank correlation suggests that a south-easterly bearing is related to a lower b value.

The parameter k(quick) is correlated with a number of characteristics (Fig. 5.10), BFIHOST (o.: -0.66, l.: -0.64), DPLBAR (o.: 0.72, l.: 0.85), the mean drainage path slope index DPSBAR (o.: -0.74, l.: -0.71), PROPWET (o.: 0.74, l.: 0.75) and average annual rainfall over a selected period SAAR (o.: -0.59, l.: -0.56). The largest correlation is found with DPLBAR (on log-scale). The variable DPLBAR is the mean drainage path distance of all nodes of the IHDTM to the catchment outlet. The result suggests that larger and more elongated catchments drain more slowly. On the contrary, steeper catchments produce smaller residence times, the mean drainage path slope (DPSBAR) is higher. The correlation with PROPWET is probably of little explanatory value. As noted above, the values of the catchments used vary between 0.33 and 0.36, the UK wide variation is between 0.20 and 0.85. The catchments tested are therefore very similar and further analysis is required.

Sefton and Howarth (1998) related the quick response residence time (within the IHACRES model structure!) to catchment size and stream-frequency, but found no improvement when including channel slope. In the study by Lamb *et al.* (2000), this parameter showed the highest correlation of the four PDM parameters tested (the fast/slow split was fixed). They derived it from BFIHOST, stream network centroid, MORECS residual soil moisture and suburban area.

The slow flow residence time k(slow) does not show any significant (5% level) correlation with any of the available catchment characteristics. This is not unexpected since this is the least identifiable parameter. The problem of identifying this parameter (and other parameters related to the low flow periods for that matter) using an OF based on the complete hydrograph is one of the main reasons why segmentation schemes were introduced (e.g., Dunne, 1999; Boyle *et al.*, 2000; Wagener *et al.*, 2001; see also Chapter 4). It seems quite unlikely that a parameter such as the slow flow residence time can be regionalized based on a calibration using an OF that emphasises the fit to high flow, such as the RMSE.

Similar results have been reported by other researchers. Lamb *et al.* (2000) derive a regional equation to estimate this parameter from BFIHOST, soil porosity and underlying geology. However, the coefficient of determination produced by this model is 0.6 and thus the lowest of their four regionalized model parameters. Their result is similar to the one by Sefton and Howarth (1998) who relate a slow flow residence time to different soil variables. The correlation coefficient between regionalized and locally calibrated values is only 0.37, however.

The fraction of effective rainfall that contributes to the quick response, alpha, is highly correlated to two catchment characteristics, the baseflow index BFIHOST and PROPWET. The first correlation (on original scale) is negative (o.: -0.91; l.: -0.96), while the second one is positive (o.: 0.84; l.: 0.89). Both are significant at the 1% level. The first correlation is obvious, while the second indicates that a wet catchment, probably containing more saturated and therefore contributing areas, produces a high percentage of quick response. Both correlations seem sensible which makes the result reassuring. Strangely, wetter catchments (higher PROPWET values) also produce higher quickflow residence times. However, concerns regarding PROPWET are outlined above.

Sefton and Howarth (1998) found the contribution to slow response (1-alpha) was highly correlated to the percentage aquifer in a catchment (0.77), a variable not available here. This split parameter is sometimes assumed to be directly equal to the standard percentage runoff (SPRHOST) or BFIHOST, e.g., Lamb *et al.* (1999) and Young (2001). SPRHOST is highly correlated to BFIHOST as shown earlier (-0.98). The result found here therefore suggests that fixing this parameter *a priori* can be justifiable.

5.5.4.2 Effect of using FH and FL

Using the OFs, FH and FL, leads to some very interesting changes in parameter–catchment characteristic correlation compared to using the RMSE.

The maximum storage capacity, which formerly only showed insignificant correlations, becomes correlated at the 5% level to BFIHOST for FH (0.71), and to ASPVAR (invariability in slope aspect) for FL (-0.70), both at original scale.

Parameter b is correlated at the 1% level to DPSBAR (o.: 0.84, l.: 0.88) and to SAAR (l.: 0.89) when using FH. The result therefore indicates a larger variation in (model) storage capacities with greater drainage path slopes and higher rainfall. No significant correlation at the 5% level is found for low flows, FL.

The result for the quickflow residence time k(quick) is probably the most interesting. For FH, this parameter is highly correlated to catchment size and drainage path configuration (o.: 0.84, significant at the 1% level). However, for small runoff events during dry periods analysed using FL, k(quick) is highly correlated to the rainfall characteristics represented by SAAR (o.: 0.96, l.: 0.99) and RMED-2D (o.: 0.92, l.:0.90).

The parameter k(slow) is also correlated to different variables depending on the OF used. Applying FH, k(slow) is positively correlated to $URBEXT_{1990}$ (o.: 0.87, l.: 0.91), and negatively to the variability in aspect (ASPVAR, o.: -0.77, l.: -0.75) and RMED-2D (o.: -0.74, l.: -0.77). However, the purpose of this parameter within the local model structure is to fit the long recession periods, which will usually be below the mean and therefore analysed by FL. The optimum k(slow) values with respect

to FL show a high correlation to BFIHOST (o.: 0.80, l.: 0.80) and PROPWET (o.: -0.83, l.: -0.79). This result (compared with the lack of correlation when using RMSE) suggests that a regional model for this parameter is probably better derived from the local model fit to low flow periods.

The quickflow contribution alpha shows very similar correlations for high and low flows with slightly higher correlation values for FH. This is probably caused by the high identifiability of this parameter at the end of high flow events (when there is a change between quickflow dominated and baseflow dominated periods) as demonstrated using the dynamic analysis in Chapter 4.

5.5.4.3 Regional structures and performance criteria

Different regional model structures are derived based on the results of the correlation analysis and subsequently compared. All structures assume independence between the dependent and independent variables.

The following functional forms were tested,

$$\theta_L = \theta_{R0} + \theta_{R1}\phi_1 + \theta_{R2}\phi_2 + \ldots \tag{5.23}$$

where θ_L is the dependent variable (local model parameter) and ϕ_i are the independent variables (catchment characteristics). The regional parameters are represented by θ_{Ri}. The second form is

$$\ln(\theta_L) = \theta_{R0} + \theta_{R1}\ln(\phi_1) + \theta_{R2}\ln(\phi_2) + \ldots \tag{5.24}$$

which, on exponentiation, results in

$$\theta_L = e^{\theta_{R0}}\phi_1^{\theta_{R1}}\phi_2^{\theta_{R2}}\ldots \tag{5.25}$$

The values for ϕ_i cannot be zero in this case. The variable URBEXT is therefore transformed to (1+URBEXT), in line with the procedure proposed by Houghton-Carr (1999, p. 241).

The regional models based on multiple, weighted and sequential regression are developed using a stepwise procedure where different combinations of independent variables were tested until the best model was found. The statistics calculated to judge the performance of the different regression models are the RMSE (Table 3.1), the coefficient of determination (R^2); (the amount of variability explained by the regression

model), the overall F statistic for the regression, and its probability p. The coefficient of determination can be calculated as follows (Kottegoda and Rosso, 1998, p. 371)

$$R^2 = \frac{SS_R}{SS_{yy}} \tag{5.26}$$

"This is the ratio of the sum of squares due to regression (SS_R) to the total sum of squares (SS_{YY})" (Kottegoda and Rosso, 1998, p.371). These sums of squares can be derived using (Kottegoda & Rosso, 1998, p. 366)

$$SS_R = \hat{\theta}_R{}^T \Phi^T \theta_L - \frac{\left(\sum_{i=1}^{n} \theta_{Li}\right)^2}{n} \tag{5.27}$$

and

$$SS_{yy} = \theta_L{}^T \theta_L - \frac{\left(\sum_{i=1}^{n} \theta_{Li}\right)^2}{n} \tag{5.28}$$

where θ_L is the vector of observations, n is the number of observations, Φ is a matrix of regressors and $\hat{\theta}_R$ is a vector of (estimated) regression parameters.

The coefficient of determination describes how much of the total data variance can be described by the model (Legates and McCabe, 1999). It can take values between 0 and 1, with 1 being optimum.

The R^2 statistic is listed in Table 5.10 together with the regional models derived. The remaining statistics are not shown.

5.5.4.4 Multiple univariate regression – benchmark

The most obvious regional model is the one for the parameter alpha (R^2 = 0.91), using BFIHOST as regressor (Table 5.10[a]). This is in accordance with the result of the correlation analysis which showed an extremely high value for the relationship between alpha and BFIHOST. Including PROPWET in the regional model leads to a regional parameter not significantly different from zero.

The second best regression model is the one for k(quick), which can be predicted using DPLBAR. However, the performance (R^2 = 0.72) is less than for alpha. The scatter plot in Fig. 5.10 (bottom right) suggests that there is an outlier (the smallest k(quick) value from Medina@ ShideWeir) that might prevent a better correlation between the k(quick) values and DPSBAR. According to the discussion held before, it might be sensible to ignore this value during the analysis and accept that this parameter might not be explainable for some catchments using the available information (*i.e.,* catchment characteristics). The resulting regional model, when ignoring the outlier, is

$$k(quick) = 12.5824 \cdot DPSBAR^{-0.3497} \tag{5.29}$$

which has a better R^2 (0.90) than the first equation (0.72) and also the remaining statistics are better. However, this lack of robustness is probably again due to the small number of data points.

The parameter b can also be predicted using DPLBAR (and the full set of catchment characteristics), but with a very low performance (R^2 = 0.44) and with one regional parameter not significantly different from zero. Adding BFIHOST to the equation slightly improves the R^2 value, but the remaining statistics degrade.

Three catchment characteristics are required to yield a useful model for cmax, *i.e.,* BFIHOST, DPSBAR and ASPBAR. The performance of the regional model is reasonable, with an R^2 value of 0.67. However, none of the regional parameters is significantly different from zero. Using the median of all locally calibrated values might be a more reliable model in this case.

No regional model could be derived for the slow flow residence time k(slow). The median value of all local k(slow) values is therefore used as the regional model following Seibert (1999b).

5.5.4.5 Multiple weighted univariate regression – considering identifiability

Figure 5.11 shows the maximum identifiability values (using the measure introduced in Chapter 4) for the pd4-2pll parameters in all of the 10

Table 5.9 Correlation coefficients for all variables involved. Local model parameters are optimal with respect to the two OFs FH and FL. Three catchments are excluded from this calculation: (1) Teise@Stonebridge, because it is kept as a test catchment; (2) Test@Broadlands, due to the differences in BFIHOST; and (3) Blackwater@Ower, because no satisfactory calibration could be established. (The best parameter values are normally distributed, but show individual outliers. Normality plots are not shown.)

(a) Correlation coefficients for variables on original scale using FH.

Characteristic	cmax	b	k(quick)	k(slow)	alpha
BFIHOST	0.7052*	0.4587	-0.3720	0.0077	
DPLBAR	-0.0958	-0.4448		-0.0320	0.6997*
DPSBAR	-0.3011		-0.2190	-0.3150	-0.5372
PROPWET	-0.6178	-0.5246	0.4553	-0.0391	
SAAR	-0.0421		-0.1688	-0.5490	-0.6856*
$URBEXT_{1990}$	0.4675	-0.5602	0.0645		0.2869
ASPBAR	-0.4263	-0.2797	0.0721	-0.6061	0.3440
ASPVAR	-0.3644	0.1824	-0.5643	-0.7703*	-0.2951
RMED-2D	-0.2981	0.5290	0.0881	-0.7356*	-0.2976

(b) Correlation coefficients for variables on log-transformed scale using FH.

Characteristic	cmax	b	k(quick)	k(slow)	alpha
BFIHOST	0.5822	0.4824	-0.3659	0.0145	
DPLBAR	-0.0969	-0.4544		-0.0224	0.7354*
DPSBAR	-0.3818		-0.1487	-0.3740	-0.5378
PROPWET	-0.4457	-0.5536	0.4763	-0.0286	0.9206**
SAAR	-0.1816		-0.1203	-0.5862	-0.6559
$URBEXT_{1990}$	0.5580	-0.5691	0.0235	0.9106**	0.2387
ASPBAR	-0.3465	-0.4044	0.0715	-0.6192	0.4653
ASPVAR	-0.3799	0.2766	-0.5780	-0.7499*	-0.2905
RMED-2D	-0.3581	0.6222	0.1427	-0.7663*	-0.2432

Table 5.9 (cont.)

(c) Correlation coefficients for variables on original scale using FL.

Characteristic	cmax	b	k(quick)	k(slow)	alpha
BFIHOST	-0.4734	0.5717	0.4539	[illegible]	-0.7763*
DPLBAR	0.3771	-0.5449	-0.0818	-0.4877	0.3237
DPSBAR	-0.3582	0.2320	0.6266	0.1819	-0.2858
PROPWET	0.4908	-0.6141	-0.3728	[illegible]	0.7405*
SAAR	-0.5314	0.2413	0.9624**	0.2792	-0.4778
$URBEXT_{1990}$	0.6405	-0.0946	-0.7559*	0.0954	0.1493
ASPBAR	-0.2577	-0.1355	0.0638	-0.4891	0.2462
ASPVAR	-0.6979*	0.3168	0.3471	-0.0908	-0.1002
RMED-2D	-0.4302	0.0052	0.9232**	-0.1501	-0.2391

(d) Correlation coefficients for variables on log-transformed scale using FL.

Characteristic	cmax	b	k(quick)	k(slow)	alpha
BFIHOST	-0.4698	0.5820	0.4745	[illegible]	-0.7289*
DPLBAR	0.5300	-0.6777	-0.1765	-0.5137	0.6051
DPSBAR	-0.3622	0.4582	0.7447*	0.1190	-0.2864
PROPWET	0.5112	-0.6358	-0.4133	-0.7848*	0.7189*
SAAR	-0.4998	0.4840	0.9850**	0.3052	-0.3948
$URBEXT_{1990}$	0.5738	-0.3080	-0.7187*	0.0631	0.1196
ASPBAR	-0.3081	0.0570	-0.0740	-0.5371	0.0196
ASPVAR	-0.6749	0.3886	0.3731	0.1246	-0.2115
RMED-2D	-0.3841	0.2611	0.9042**	-0.0666	-0.1550

*Significant at the 5% level.
**Significant at the 1% level.

catchments. The catchments are sorted according to increasing BFIHOST values, left to right (see Table 5.6 for indices). Some interesting trends are revealed in this plot.

Parameter b appears to be most identifiable in mixed response catchments, *i.e.*, those with medium BFIHOST values. It is less identifiable towards the extreme ends where the response is either very flashy (low BFIHOST value) or very damped.

The identifiability of k(quick) reduces with increasing BFIHOST values. This seems sensible, because it can be expected that the importance of the quickflow routing component in reproducing the overall response (as measured by RMSE) also reduces in a similar way.

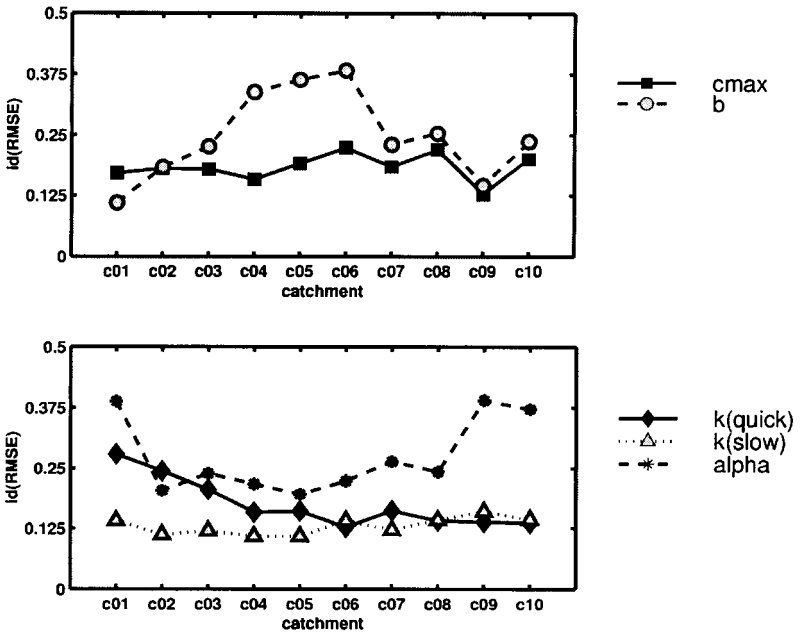


Fig. 5.11 Variation of parameter identifiability with catchment. The catchments are sorted according to increasing BFIHOST values (Table 5.6).

The remaining two parameters, k(slow) and cmax, do not show any obvious trend, apart from a general lack of identifiability, in particular with respect to k(slow).

These differences in identifiability can also be seen as an indication of the importance of individual components to represent a particular catchment (within the limitation of using a single OF). The identifiability values are used as weights in the weighted regression procedure outlined in Section 5.3.3. The resulting regional models are listed in Table 5.10(b). It can be seen that the performance (measured using R^2) of the regional models for parameters k(quick) and alpha increases from 0.72 to 0.94, and from 0.91 to 0.97, while the performance for the parameters cmax and b is only slightly reduced (0.67 to 0.63 and 0.44 to 0.43). Additionally, it is now possible to derive a regional model for k(slow), but with very wide confidence limits on the regional parameter.

Table 5.10 The regional models resulting from the different approaches: (a) Conventional regression as a benchmark, (b) weighted regression, and (c) sequential regression. Parameters for which no sensible relationship could be found are set to their median values.

(a) Conventional regression as benchmark.

Local parameter	Regional model	Lower CFL*	Regional parameter	Upper CFL*	R^2
cmax	$606.57 + 224.94 \cdot BFIHOST$ $-4.28 \cdot DPSBAR - 1.17 \cdot ASPBAR$	-128.01	606.57	1341.16	0.67
		-830.3	224.94	1280.0	
		-14.74	-4.2808	6.176	
		-4.982	-1.1711	2.639	
b	$3.1680 \cdot DPLBAR^{-1.1014}$	0.0250	3.1680	192.944	0.44
		-2.5368	-1.1014	0.3340	
k(quick)	$1.5391 \cdot DPLBAR^{0.2034}$	1.0179	1.5391	2.3270	0.72
		0.0590	0.2034	0.3479	
k(slow)	190.43 (median)	-	-	-	-
alpha	$0.2893 \cdot BFIHOST^{-1.1258}$	0.2197	0.2893	0.3809	0.91
		-1.5259	-1.1258	-0.7257	

*90% Confidence Limits

(b) Weighted regression.

Local parameter	Regional model	Lower CFL*	Regional parameter	Upper CFL*	R^2
cmax	$610.43 + 241.70 \cdot BFIHOST$ $-4.39 \cdot DPSBAR - 1.18 \cdot ASPBAR$	-111.90	610.43	1322.77	0.63
		-439.47	241.70	922.86	
		-11.119	-4.3858	2.3474	
		-3.5584	-1.1806	1.1972	
b	$1.4024 \cdot DPLBAR^{-0.8399}$	0.0108	1.4024	181.544	0.43
		-2.4733	-0.8399	0.7935	
k(quick)	$1.5278 \cdot DPLBAR^{0.2050}$	1.0089	1.5278	2.3131	0.94
		0.0593	0.2050	0.3507	
k(slow)	$0.2319 \cdot PROPWET^{-6.3269}$	$e^{-18.0443}$	0.2319	$e^{15.1216}$	0.57
		-21.777	-6.6146	9.1234	
alpha	$0.2799 \cdot BFIHOST^{-1.1953}$	0.2130	0.2799	0.3678	0.97
		-1.5923	-1.1953	-0.7982	

*90% Confidence Limits

Table 5.10 (cont.)

(c) Sequential regression.

Local parameter	Regional model	Lower CFL*	Regional parameter	Upper CFL*	R^2
cmax	$e^{8.4483} \cdot DPSBAR^{-0.9686} \cdot ASPVAR^{-0.5217}$	$e^{0.0153}$ -2.7490 -1.8439	$e^{8.4483}$ -0.9686 -0.5217	$e^{16.8813}$ 0.8118 0.8006	0.48
b	0.1624 (median)	-	-	-	-
k(quick)	$1.7458 \cdot ASPVAR^{-0.1899}$	1.0357 -0.3955	1.7458 -0.1899	2.9435 0.0157	0.53
k(slow)	197.87 (median)	-	-	-	-
alpha	$0.2893 \cdot BFIHOST^{-1.1258}$	0.2197 -1.5259	0.2893 -1.1258	0.3809 -0.7257	0.91

*90% Confidence Limits

Analysing the conventional and weighted regional models in detail shows that the models for cmax, k(quick) and alpha are very similar. The confidence limits for the regional parameters of the cmax model are smaller, though. The regional model to predict b is different; whether it performs better in the test catchment will be seen later.

5.5.4.6 Generalized regression

It was already mentioned that the only parameter for which the Spearman rank correlation coefficient is considerably higher than the linear correlation is b (Table 5.8). This suggests a possible non-linear relationship that probably cannot be described by the conventional regression model structures applied here.

A GRNN is therefore used to derive a regional model with respect to this parameter using ASPBAR, *i.e.,* the dominant aspect of catchment slopes. The performance of a GRNN cannot be assessed using R^2 since it can always provide a perfect fit to the data. The model will have to show its merit when applied to the test catchment in the next section.

5.5.5 Alternative regional procedure

Different variations on the traditional procedure were described earlier. Only the sequential method (Lamb *et al.*, 2000; Lamb and Calver, 2002) is analysed here since it seems the most promising way to reduce the problem of lack of identifiability without ignoring the physical (or maybe better conceptual) relevance of the model parameters.

5.5.5.1 Sequential procedure

At every iteration step, the identifiability of all (free) model parameters is analysed using the identifiability measure introduced in Chapters 3 and 4. The measure is calculated as the average of all maximum identifiability values over all catchments for each of the five parameters. This way, it is possible to see whether the assumption that the identifiability of the remaining parameter increases is correct and it allows an objective selection of the parameters to be regionalized during every iteration (in contrast to the subjective selection by Lamb *et al.*, 2000). It was hoped that initially unidentifiable parameters might become identifiable due to the sequential reduction in model complexity.

Figure 5.12 shows the average identifiability of each of the five parameters during the five iteration stages. One can easily see that the identifiability of the remaining parameters increases every time one of the parameters is fixed. Even k(slow), which initially is not identifiable at all, ultimately shows clear optima in the various catchments. However, the improvement in identifiability did not lead to improved regional relationships for the different parameters (Table 5.10).

The most identifiable parameter after the first iteration is alpha, for which the regional model is therefore identical to the one derived in the benchmark regression. This parameter is subsequently fixed to its regional value and another URS with 10,000 samples is executed for all catchments.

The shape parameter b is fixed following the second iteration. However, it was not possible to derive a sensible regional relationship to any of the catchment characteristics available. This parameter is therefore fixed to its median value of 0.1624.

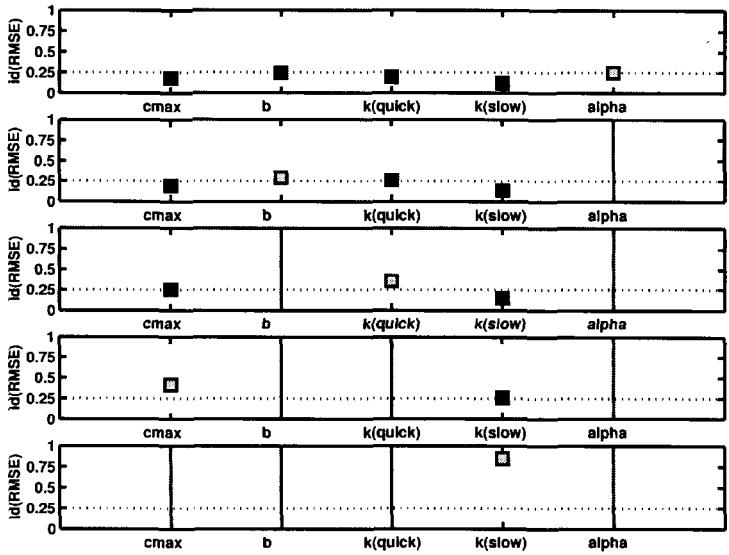


Fig. 5.12 Variation in parameter identifiability with iteration step during the sequential regionalization procedure. A vertical bar indicates a fixed parameter. The most identifiable parameter for every iteration is shown as a grey rectangle. The first iteration is shown at the top, one parameter is fixed during each iteration (from top to bottom).

The next two parameters k(quick) and cmax (after another iteration) show some correlation with available characteristics, ASPVAR, and DPSBAR and ASPVAR respectively, but their regional models yield lower R^2 values than the benchmark models. Also, the uncertainty for the intercept of the regional model for cmax is very high.

It appears that the bias, which is introduced every time when one of the parameters is fixed to its regional value, reduces the chance of finding correlations for the remaining parameters. One reason for this effect is probably the use of a single OF. Hogue *et al.* (2000) show that a local sequential calibration procedure can be very successful for the Sacramento model structure. They used two OFs, the normal RMSE (emphasizing the fit to high flows) and the RMSE of the log-transformed data (emphasizing low flows). Splitting the model parameters into two groups, each one better identifiable using one of the two OFs, gave a satisfying calibration result; the overall fit to the hydrograph was good and not biased towards a particular feature as in single-objective calibration.

A similar sequential approach utilizing different OFs might be successful at regional scale.

The variation in flow simulation performance over the iteration steps, measured using the NSE criterion, is visualized in Fig. 5.13. One can see that the only significant loss in performance is found for the Medina@ShideWeir after the second iteration during which the shape parameter b was fixed. The optimum value for this catchment is considerably higher (≈ 0.8) than for the rest. The performance degrades because b was set to the median value of 0.1624. The other degradation is found after cmax is fixed (iteration 4). The regional model for this parameter does not perform well. The remaining small fluctuations in performance are probably due to sampling errors within the Monte Carlo procedure.

It emerged during this exercise that an identifiability value of about 0.25 is similar to a parameter population that has a (subjectively) clear optimum; *i.e.,* it can be said that the parameter is identifiable.

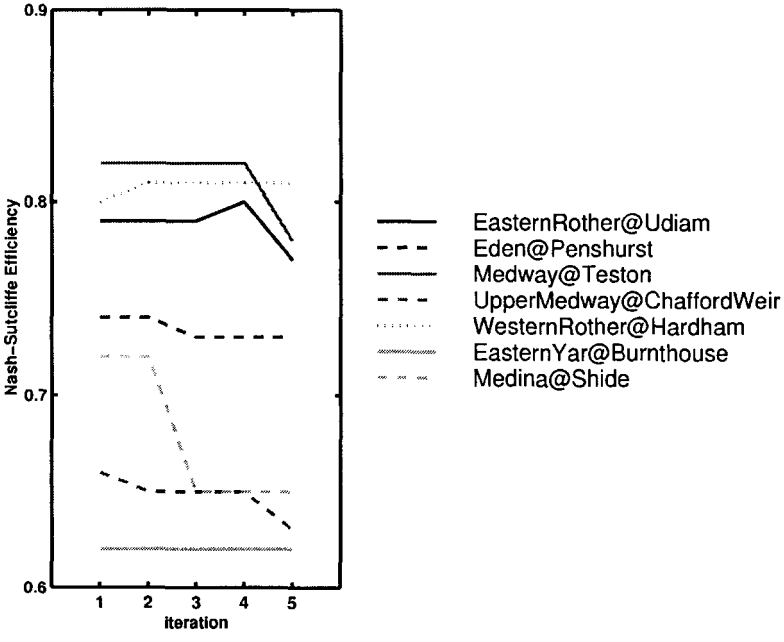


Fig. 5.13 Changing Nash-Sutcliffe Efficiency during sequential procedure.

5.5.6 Test catchment performance including flow prediction

5.5.6.1 Comparison with test catchment

The first step when analysing an ungauged catchment within the suggested framework is the calculation of the BFIHOST value in order to add the catchment to a cluster.

Boorman *et al.* (1995) derived a regional multiple regression model for the UK to derive BFIHOST values from the fractions of the different HOST classes present in a catchment. The general model has the following form (Boorman *et al.*, 1995, p.32),

$$BFIHOST = \theta_1 HOST_1 + \theta_2 HOST_2 + ... + \theta_{29} HOST_{29}$$

where $HOST_1 ... HOST_{29}$ are the fractions of each of the HOST classes, and $\theta_1 ... \theta_{29}$ are the regression parameters.

The appropriate fractions and the regression parameters for a large number of UK catchments are given in Boorman *et al.* (1995, p.33 & 86ff.). The resulting BFIHOST value is 0.416, which is sufficiently close to the FEH value of 0.443 to suggest that a classification based on this calculated value is possible. However, including the standard error on the regression parameters means that the BFIHOST value for this catchment lies between 0.366 and 0.466.

The local parameters derived using the different regional models are listed together with the values derived through local calibration in Table 5.11. The corresponding dotty plot can be found in Fig. 5.14.

Prediction of cmax using the conventional and the weighted approach are surprisingly good, considering the low performance of the regional model. The value estimated by the sequential approach is considerably lower.

The variation in b values is not large, reflecting the regional homogeneity of this parameter (except for Medina@ShideWeir, all values are between 0.01 and 0.26). The conventional approach overpredicts this parameter. The estimate using the GRNN is far too high if the outlier of

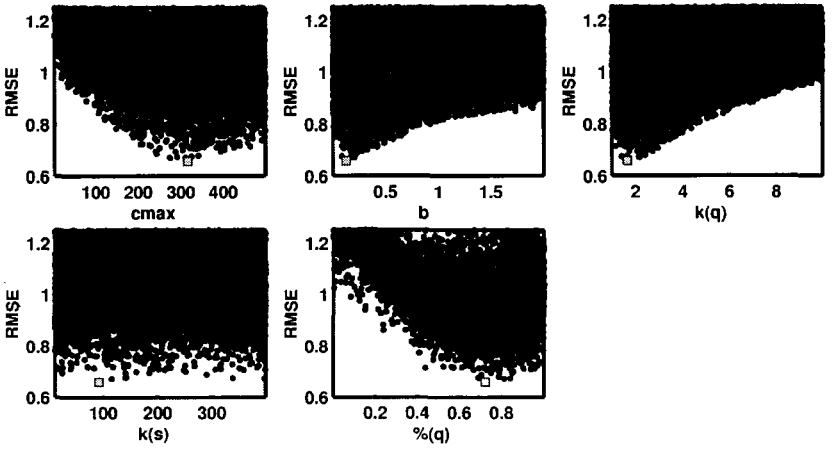


Fig. 5.14 Dotty plot for the test catchment, the Teise@Stonebridge. Result based on uniform random sampling with 10,000 parameter sets. The best parameter values are indicated by rectangles.

Medina@ShideWeir is included. Adding additional variables to ASP-BAR does not improve the model performance. Using such an ANN model can therefore only be justified through a sufficient number of successful test cases. It is unlikely to be useful within the scope of this study.

The quickflow residence time k(quick) is rather low for the test catchment. However, an analysis of the corresponding dotty plot (Fig. 5.14) shows that the density of the distribution top for this parameter is low and an optimum value might lie slightly higher (~2).

The local k(slow) value is considerably lower than all the regional estimates. No significance should be attached to this result due to the local identifiability problems described earlier.

The result for alpha is very good and therefore consistent with the quality of the regional models for all approaches.

The Nash-Sutcliffe Efficiency (NSE) values show that different parameter sets can yield a similar performance with respect to a single OF in a humid catchment. The performance only degrades for the sequential parameter set due to the cmax value, fixed at iteration 4, which is too low to give an optimum performance (Fig. 5.13).

Table 5.11 Comparison of regionalized and locally calibrated parameter values for the Teise@Stonebridge.

Approach	cmax	b	k(quick)	k(slow)	alpha	NSE
Local calibration	316.5	0.1246	1.66	92.1	0.724	0.79
Conventional	304.4	0.1810	2.61	190.4	0.724	0.78
Weighted	306.5	0.1581	2.60	148.8	0.741	0.78
Sequential	227.2	0.1624	2.70	197.9	0.724	0.76
Generalized		0.9128 (0.1319*)				

*Excluding the result for Medina@Shide. Result does not improve when using additional catchment characteristics beyond ASPBAR.

5.5.6.2 Flow prediction

The derived parameter estimates are actually more interesting than the estimated flow hydrograph from a scientific point of view. It was already stated that many parameter sets could produce a reasonable flow estimate in wet environments such as the UK. Analysing the question of how far the regionalized parameters deviate from those derived by local calibration has therefore a higher priority. Nevertheless, the performance of the different regionalized parameter estimates in reproducing the ungauged catchment response is examined here.

Additionally, one would like to assess the uncertainty in the predicted flow values due to the uncertainty in the regionalization procedure. However, the inference of statistical properties of the derived regional models is indicative at best due to the small number of catchments available. Only a point prediction of flow is therefore given here.

The differences in flow prediction are small and similar to the differences in the parameter values (Fig. 5.15). This result is probably enhanced by the fact that the Teise@Stonebridge lies relatively in the middle of the catchment clustering with respect to BFIHOST. The plot in Fig. 5.15 shows a high flow period on original scale, while the bottom plot is a low flow period on log-transformed scale.

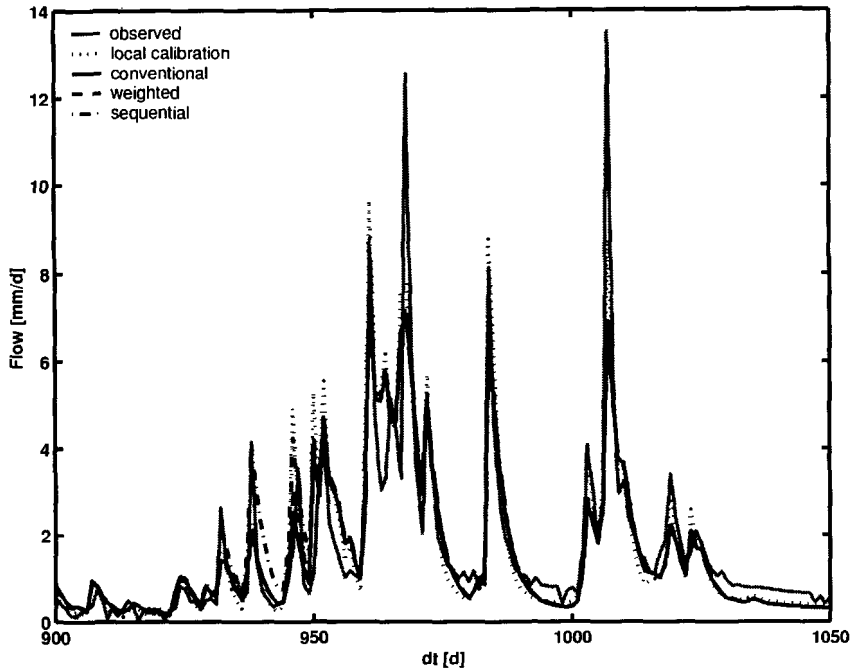


Fig. 5.15 Flow prediction for the Teise @ Stonebridge using the different regionalized parameter values and the local parameter set that is optimum with respect to the RMSE. The top plot shows a high flow period on original scale, while the bottom plot is a low flow period on log-transformed scale.

5.6 Summary and Conclusions

The following observations with respect to regional modelling were made:

- Many regionalization studies assume that a single general model structure exists, suitable for all catchments analysed. The variation in response behaviour between catchments could then be considered through different parameter sets. This is not necessarily appropriate if the catchments' characteristics vary considerably. A variable describing the geology of a catchment might be suitable as a clustering criterion in humid regions. The baseflow

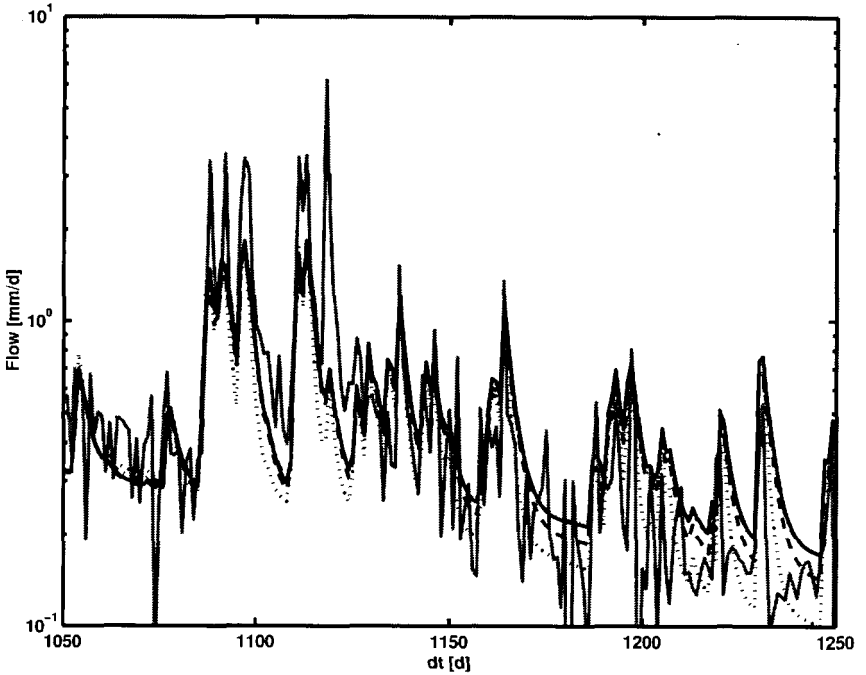


Fig. 5.15 (cont.)

index (BFIHOST) is suggested for this purpose here. However, only one catchment is singled out for testing in the application example in this chapter. A larger number of catchments is required for further testing of the suggested clustering approach.

- A wide range of regional (regression) model structures can be applied. Weighted regression, using the earlier introduced identifiability measure as a weight, seems a sensible way to include local parameter uncertainty. It has been shown that the identifiability of individual parameters varies widely between ten different catchments in South East England. Giving more importance to more identifiable parameters using the weighted least squares regression approach might lead to more robust regional models.

- Uncertainty in flow prediction in the ungauged catchment can be considered by estimating the standard uncertainty around the regional parameter estimates. A Monte Carlo sampling approach can then be used to estimate local parameters for the ungauged catchment. Local parameter values outside the feasible parameter ranges (e.g., below zero) can be given an *a priori* likelihood of zero.
- The limited application example utilizing only ten catchments showed a very high correlation between the fraction of quickflow contribution and the BFIHOST. Quickflow residence time was highly correlated to an index describing catchment size and drainage path configuration.
- Applying two different OFs (one measuring high flow fit and one low flow fit) to a single model structure showed that the resulting optimum local parameter values and therefore the resulting correlations with catchment characteristics vary widely. For example, the slow flow residence time of a parallel routing structure only showed a significant correlation if it was estimated using the low flow OF. The quick flow residence for high flows was correlated to topographic catchment features. In contrast, it was highly correlated to rainfall characteristics when only small runoff events during dry periods were analysed. Considering the above-mentioned trade-off between high and low flows, this result suggests a dependency of the regional models on the anticipated modelling purpose.
- A comparison of different metric and parametric SMA model components, all connected to identical parallel routing modules, showed that some parameters seem to be dependent on the model structure in which they are used. For example, the quickflow contribution of the metric model structure was always lower or equal to the other two parametric structures for all ten catchments.
- A sequential regionalization procedure was tested in which one local model parameter was regionalized and then subsequently set to the regional value while the model was recalibrated to all catchments. The identifiability of the remaining parameter in-

creased with every iteration. However, the bias in fit introduced in this procedure is thought to be the main reason that the regional relationships deteriorated compared to the conventional approach.

- A new framework for regional modelling has been suggested based on the review in this chapter and the findings of the application example.

Chapter 6

Discussion, Conclusions and Recommendations for Future Research

Il est vain, si l'on plante un chêne, d'espérer s'abriter bientôt sous son feuillage.

Antoine de Saint-Exupéry (1900–1944), *Terre des Hommes*

6.1 Discussion and Conclusions

6.1.1 General

Two fundamental questions in lumped conceptual rainfall-runoff modelling have still not been satisfactorily answered in forty years of research effort:

- What is the appropriate model structure for a given (type of) hydrological system and a particular modelling task?
- What is the appropriate parameter set within this structure to characterize the unique response features of a particular catchment?

An extensive literature survey (Chapters 1 and 2) magnified the above stated problems and highlighted some suggestions for potential ways forward.

Little objective guidance is available for the selection of an appropriate *model structure*. Comparison studies are often based on a single performance measure, include only a limited range of contestants (model structures), and their results are difficult to generalize. However, despite this, some general observations can be made.

- Simple structures (in terms of number of free parameters) perform as well as complex ones for many purposes.
- The number of identifiable parameters, at least with respect to a single OF, is between three and five.
- Many model structures have been developed, but only a limited number of components are used within them.

Local modelling procedures (applicable to model gauged catchments) were initially based on the assumption that hydrological models could be treated similarly to regression models, *i.e.,* simple optimization with respect to a single OF was considered sufficient. Recent research has questioned the usefulness of this paradigm and some researchers call for more *conceptual* modelling approaches that better recognize the nature of the model structures used. The lack of parameter identifiability has led some researchers to the conclusion that the idea of an optimum parameter set should be abandoned completely in favour of a population of acceptable parameter sets (or models). This explicitly points to uncertainty which has to be acknowledged. Additionally, the influence of model structural error appears to be higher than previously recognized. However, no suitable approach to explicitly assess this uncertainty is currently available.

Some applications of *regional procedures* (applicable to model ungauged catchments) for conceptual model parameter estimation can be found in the literature, but few analyse the elements and possible alternatives of those procedures in detail. Neither have the presence and influence of uncertainty in regionalization been given appropriate consideration.

A three-step approach has been used to address the three areas reviewed above:

- Develop a rainfall-runoff modelling and analysis toolkit to implement and evaluate different model structures.
- Review and improve existing procedures for local modelling (modelling of gauged catchments).
- Review and evaluate the idea of parameter regionalization (regional modelling or modelling of ungauged catchments).

6.1.2 Discussion and conclusions with respect to the toolkit

A toolkit that facilitates the development and analysis of lumped and parsimonious model structures has been implemented using state-of-the-art modelling techniques (Chapter 3).

The RRMT allows the implementation of parametric, or hybrid metric-parametric model structures. Its major advantage is a high degree of structural flexibility which allows quick implementation and evaluation of different model structures to identify those suitable for the circumstances under consideration. The option to run the RRMT in batch mode is particularly attractive for studies that include a range of model structures or a large number of catchments.

The MCAT enables detailed investigation of model performance, parameter sensitivity and identifiability, model structure suitability, and prediction uncertainty. It is generic and can be applied to analyse any dynamic mathematical model structure. Within the MCAT, a new empirical identifiability measure has been implemented, based on the gradient of the cumulative parameter distribution.

A brief application example has been used to show how the toolkit can be applied. The application of a variety of model structures with different levels of complexity showed that model performance increased with increasing complexity, while parameter identifiability decreased. This demonstrated a clear trade-off between these two features. The study used an MO approach, assessing different response periods, to ensure that any improvement in performance did not go unnoticed.

6.1.3 Discussion and conclusions with respect to local modelling

The current idea of local CRR model identification, *i.e.*, the modelling of gauged catchments, has been analysed in detail (Chapter 4). It has been suggested in this monograph that the idea of establishing a 'true' model structure is ill-founded and should be replaced by an evaluation of model structures with respect to

- performance,
- (parametric) uncertainty, and
- underlying assumptions.

Current automatic methods of parameter identification also lack the complexity of subjective manual techniques and are often too simplistic in their approach. A multi-objective approach to model structure analysis has been developed to address aspects 1) and 2) of the three-dimensional evaluation described above. The third point is assessed through a novel dynamic identifiability approach (DYNIA) which analyses how (initially uniform) parameter distributions are conditioned over different data periods. A violation of underlying assumptions occurs if, for example, peaks in the response surface of a particular parameter vary in time and in parameter space. These approaches are combined in a framework of corroboration and rejection more in line with scientific methods, as for example propagated by Popper (2000), than traditional procedures for rainfall-runoff modelling.

The following main observations were made in the application examples in Chapter 4:

- Example 1 led to the conclusion that accepting the multi-objective nature of model calibration and integrating it into the modelling process increases the amount of information retrieved from the model residuals to (1) find the parameter population necessary to fit all aspects of the observed output time-series (albeit separately), (2) increase the identifiability of the model parameters, and (3) assess the suitability of the model structure to represent the natural system (*i.e.,* identify model structural insufficiencies).
- One approach to derive MOs for single output models is the segmentation of the hydrograph. A segmentation based on the slope of the observed streamflow and additional thresholds has shown to be effective in separating different response modes for humid catchments (Chapter 4, examples 1 and 3), while an approach using rainfall to find "driven" periods is difficult to apply in these areas due to the high frequency of rainfall events (see example in Chapter 3).
- A distinct difference in optimum parameter sets to fit high and low flows was found for all structures analysed in Chapter 4. This seems to be the main problem in currently available CRR model structures.

- The identifiability measure introduced in Chapter 3 is a useful empirical tool to assess individual parameters in an objective manner without having to make limiting assumptions (examples 1 and 3). Using this measure to compare the identifiability of different parameters requires the assumption that their initial feasible ranges are equally well defined.
- Applying this identifiability measure in a dynamic fashion leads to DYNIA. The approach is of general utility and simple to implement. It can be applied to (1) analyse individual model structures with respect to structural inadequacies; (2) identify suitable parameters in an approach more in line with sophisticated manual approaches; (3) separate periods of high and low information content for individual parameters, for example as a first step in a multi-objective analysis; (4) relate model parameters, and therefore model components, and response modes of the natural system; (5) investigate data outliers or anomalies; and (6) analyse the potential of experimental design, e.g., tracer experiments, to identify individual parameters.
- Some of the main results of the DYNIA applications in examples 2 and 3 were: (1) For the Penman SMA component (ic1): a) the root constant is mainly identifiable during wetting up periods; b) the bypass parameter is defined by major rainfall events during dry periods; c) the optimum bypass values change with time, indicating that a dynamic aspect is missing in the model structure; d) the correlation between root constant and bypass is reduced considerably if only periods of high identifiability are considered. (2) For the two parallel linear routing component (2pll): a) the quick and slow flow residence times are mainly identifiable during periods of quick and slow recession respectively; and b) the split parameter alpha is most identifiable during the change between quick and slow flow dominated periods after large runoff events. (3) For the probability-distributed SMA (pd3 and pd4): a) the pd3 (ae at potential rate) performs better than the pd4 (ae linearly dependent on soil moisture content) in an MO analysis; b) however, the pd4 module shows more consistent op-

timum parameter values in a DYNIA, indicating that it is the better model structure.

6.1.4 Discussion and conclusions with respect to regional modelling

The following observations with respect to regional modelling were made during a review of the idea of regionalization and the application of different approaches in a test case:

- Many regionalization studies assume that a single general model structure exists, suitable for all catchments analysed. The variation in response behaviour between catchments could then be considered through different parameter sets. This is not necessarily the case if the catchments vary considerably in hydrological response. A clustering of catchments (located in southeast England) based on the baseflow index (BFIHOST) is suggested to consider differences in underlying geology that mainly influences the model structure selection.
- A wide range of regional (regression) model structures can be applied. Weighted regression, using the earlier introduced identifiability measure as a weight, seems a sensible way to include local parameter uncertainty. It has been shown that the identifiability of individual parameters varies widely between ten different catchments in South East England. Giving more importance to better identifiable parameters using the weighted least squares regression approach might lead to more robust regional models.
- Uncertainty in flow prediction in the ungauged catchment can be considered by estimating the standard uncertainty around the regional parameter estimates. A Monte Carlo sampling approach can then be used to estimate local parameter values for the ungauged catchment. Local parameter values outside the feasible parameter ranges (e.g., below zero) can be given an *a priori* likelihood of zero.
- The limited application example utilizing only ten catchments showed a very high correlation between the fraction of quickflow contribution and the BFIHOST. Quickflow residence time

was highly correlated to an index describing catchment size and drainage path configuration.

- Applying two different OFs (one measuring high flow fit and one low flow fit) to a single model structure showed that the resulting optimum local parameter values and therefore the resulting correlations with catchment characteristics vary widely. For example, the slow flow residence time of a parallel routing structure only showed a significant correlation if it was estimated using the low flow objective function. The quick flow residence time for high flows was correlated with topographic catchment features. In contrast, it was highly correlated to rainfall characteristics when only small runoff events during dry periods were analysed. Considering the above-mentioned trade-off between high and low flows, this result suggests a dependency of the regional models on the anticipated modelling purpose.
- A comparison of different metric and parametric SMA model components, all connected to identical parallel routing modules, showed that some parameters appear dependent on the model structure in which they are used. For example, the quickflow contribution of the metric model structure was always lower or equal to the other two parametric structures for all ten catchments.
- A sequential regionalization procedure was tested in which one local model parameter was regionalized and then subsequently set to the regional value while the model was recalibrated to all catchments. The identifiability of the remaining parameter increased after every iteration. However, the bias in fit introduced in this procedure is thought to be the main cause of deterioration of the regional relationships compared to the conventional approach.
- A new framework for regional modelling has been suggested based on the review in Chapter 5 and the findings of the application example.

6.2 Recommendations for Future Research

6.2.1 Recommendations with respect to the toolkit

- The idea of a component library can be taken much further. It is envisaged that future modelling systems will contain components that support the user in finding an appropriate implementation of his or her perceptual model. A large number of algorithms of varying complexity and varying types of process description could be made available in a component library. The user would add information about (1) catchment characteristics, e.g., the dominant runoff production mechanism might be infiltration excess, which excludes all algorithms assuming saturation excess; (2) available data (complex algorithms might not be applicable if only measurements of streamflow, precipitation and temperature are available); (3) the modelling purpose (a very simple baseflow component might suffice if the purpose is the modelling of a single event). This could be combined in a knowledge-based approach to rainfall-runoff modelling as suggested by Wagener (1998). The remaining algorithms could then be used in an optimization procedure to derive a ranking of model structures. The process of finding an appropriate model structure could be implemented in a way analogous to genetic programming. In genetic programming, not only the model parameters are optimized, but also their functional relationship, *i.e.*, the model structure.
- The current structure of the RRMT shows a sequential division into mainly vertical and mainly horizontal movement of water in the catchment. The first is represented by the soil moisture accounting (SMA) component, while the second is represented by the routing (R) component of each model structure. This originated from an initial desire for a simple structure, with readily interchangeable components, as for example found in the IHACRES structure, but has been shown to be of limited flexibility. The effective rainfall produced by the SMA component is

usually split into a quick and a slow part in the R component. These are then routed with two different time constants and combined to an estimate of streamflow. It would be more realistic to split the flow in the SMA component, which would also allow the flow generation to be made a function, for example, of the current moisture state of the model, followed by two separate routing components such as one for quick and one for slow flow. This is also possible within the current modelling framework, but the R component has to be integrated into the SMA component, a solution which is not very elegant.

- The SMA and routing modules in the RRMT and the DYNIA in the MCAT should be reprogrammed in C/C^{++} or Fortran in order to increase the simulation speed. The use of larger samples in the DYNIA procedure is particularly hindered at present by long run times. However, the identifiability measure used within the DYNIA approach has proved to be quite robust even when samples sizes are small.
- Semi-distributed modelling has drawn increasing interest in the last few years. The problem of parameter estimation in this approach is largely identical to the problem in regionalization. A semi-distributed version of the toolbox should be developed and some of the methods developed in this study should be applied.

6.2.2 Recommendations with respect to local modelling

- Contrasting response periods are mainly quickflow-dominated high flow periods and baseflow-dominated low flow periods. A more physically based segmentation scheme might be useful to improve the reliability of the approach used to derive multiple OFs for single output CRR models.
- The real power of the DYNIA approach could not be assessed during this study. While a range of application examples have shown the great potential of the approach, more research is needed. Particularly, more and different model structures should be analysed.

6.2.3 Recommendations with respect to regional modelling

- The regional model structures used in this study treated the local model parameters as independent entities, following common practice. This is an assumption that will not be valid in many cases; parameter dependence is ubiquitous. Multiple-objective calibration or a sequential regionalization can improve on this. However, more sophisticated regression techniques, such as Seemingly Unrelated Regression (SUR, Chapter 5) might be more sensible. Another option is the adoption of more conceptual ideas of model regionalization, such as the fuzzy mapping approach by Beven (2000b) as described below.
- Uncertainty in regionalization, stemming for example from a lack of local parameter identifiability, needs to be considered in order to derive results that are useful for decision making. A very simple (conventional) approach is adopted in this research, including assumptions of normality of distributions involved. These assumptions are unlikely to be fully met in practice. More flexible approaches are required.
- More conceptual (in contrast to statistical or physical) modelling approaches could allow for both, the consideration of uncertainty (due to the problem that a unique best model cannot be found) and the requirements for parameters to be treated as a set. Beven (2000b) suggests a fuzzy mapping of the landscape space (e.g., different unique catchments) into the model space. Intersections of areas in the model space, *i.e.*, areas of models that are behavioural for different catchments, might give a more realistic picture of similarities in catchment response behaviour. This is an interesting approach which merits further investigation.
- A basic assumption in conceptual modelling is that the local parameters have some physical interpretation. It might be possible to derive relationships with catchment characteristics based on physical reasoning for some parameters (e.g., Drane, 2000; Koren *et al.*, 2000). This option has not been explored in this study.

6.3 Resume

The last five or so years have shown that some of the assumptions underlying (particularly parametric) rainfall-runoff modelling are highly questionable. This has led some researchers to the conclusion that any progress in this field has to come by first taking a step back and rethinking the fundamental questions we try to answer and the basic assumptions we make when attempting to find a solution.

An attempt has been made to analyse the current state of rainfall-runoff model identification for gauged and ungauged catchments and to highlight weaknesses and problems. Possible solutions are investigated and alternative modelling frameworks are suggested. Short application examples provide promising results for these new approaches. However, it must be stressed that this work is largely methodological and that more extensive case studies are required. The introduction of a dynamic approach to parameter identification has opened a new dimension for model structural analysis and more research is required to define its potential. If this work inspires the creativity of other researchers to question earlier accepted paradigms and rethink basic assumptions, ultimately leading to new and improved methodologies, then its goal has been reached.

This page intentionally left blank

Appendix A

Some Investigations into Regional Modelling

Introduction

This appendix shows some initial investigations regarding regional modelling. The presented research shows some ideas which developed into the methods presented in this monograph.

Regionalization Study

Regionalization is one way to estimate model parameters in catchments without any measured runoff time series. Developing a model structure with identifiable parameters is crucial if sensible relationships to catchment characteristics are to be established. A number of model structures (API-based loss functions and linear conceptual reservoirs) of differing levels of complexity (with 7, 4 and 3 parameters) are investigated within the framework of a regionalization study. The toolboxes are used to identify a model structure that combines good performance with low model complexity, therefore maximizing parameter identifiability (Fig. A.5a). Monthly time series data for 23 USA catchments (Fig. A.5b), provided by the Model Parameter Estimation Experiment (MOPEX), are used for this purpose.

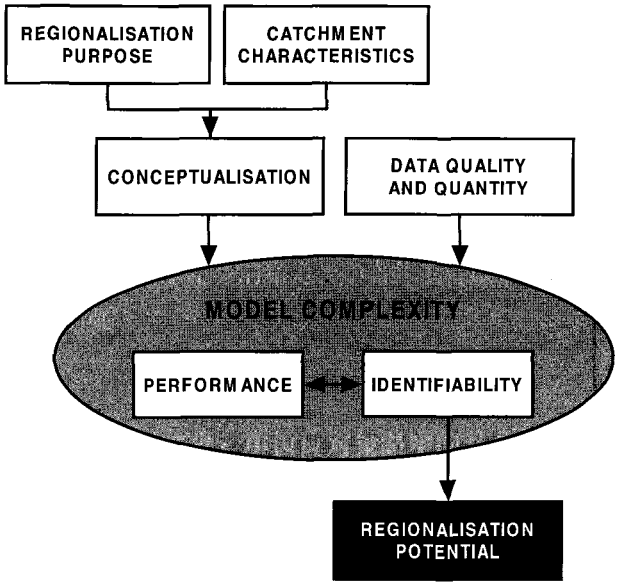


Fig. A.1 Developing a suitable model for regionalization.

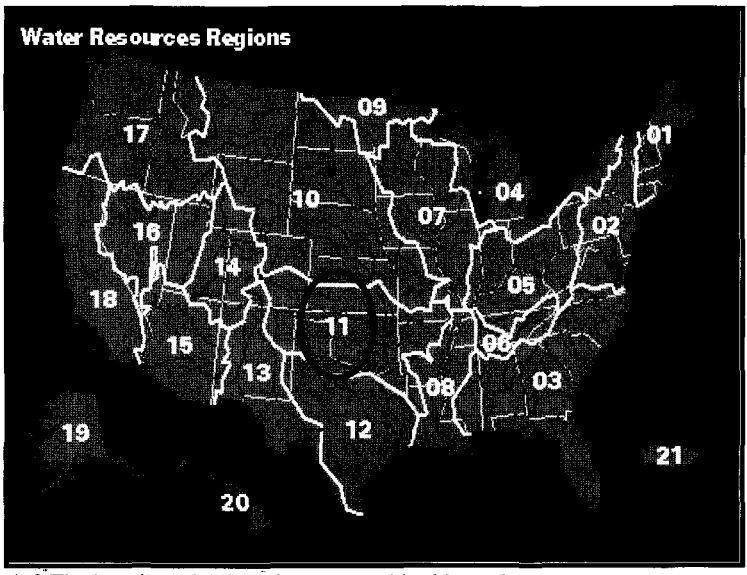


Fig. A.2 The location of the catchments used in this study.

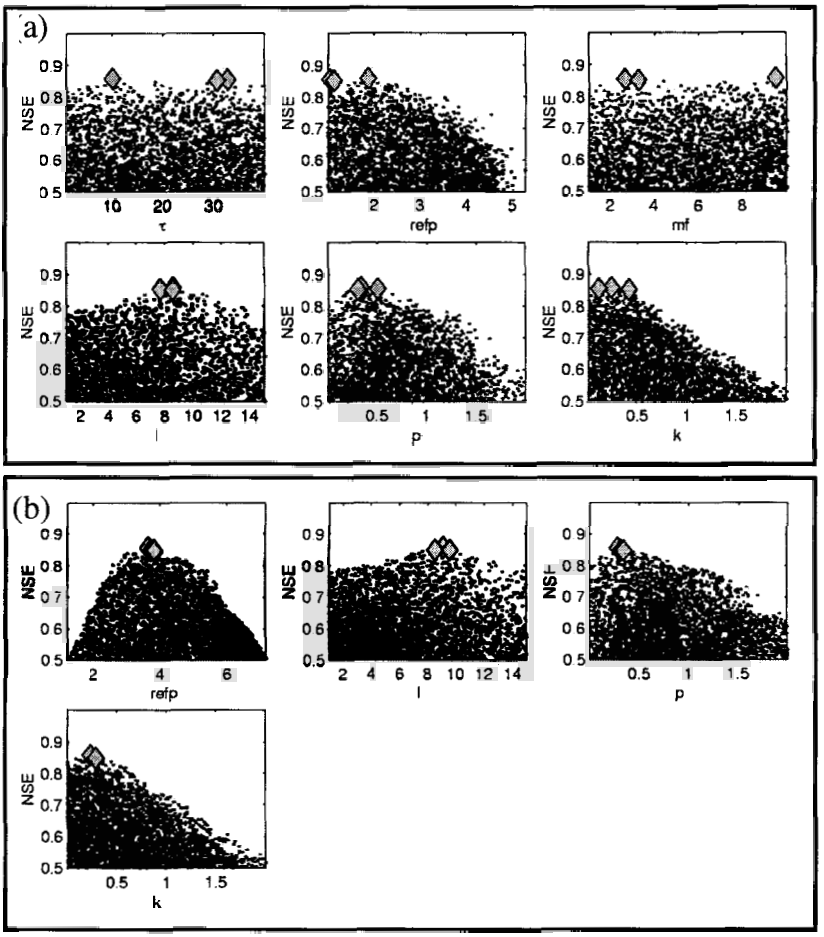


Fig. A.3 Dotty plots comparing parameter identifiability of an initial complex model (a) and a final, simplified model (b) retaining only identifiable parameters, from the regionalization example. Note that the performances of simple and complex models are similar.

A simple empirical measure (ID) is introduced to quantify the level of identifiability of different model structures (see Fig. A.6). High ID values result for parameters that are distributed over a small part of their normalised range. The average ID for all parameters gives an indication of the identifiability of a model (Fig. A.8a and A.8b).

An identifiable model (Fig. A.8b) with high prediction accuracy (Fig. A.9) is developed using the flexibility of the RRMT and the analysis methods available in MCAT (Fig. A.7).

This example shows how an appropriate model complexity, balancing identifiability and performance can be found. Further investigation is required to determine how this increased identifiability influences prediction uncertainty in ungauged catchments.

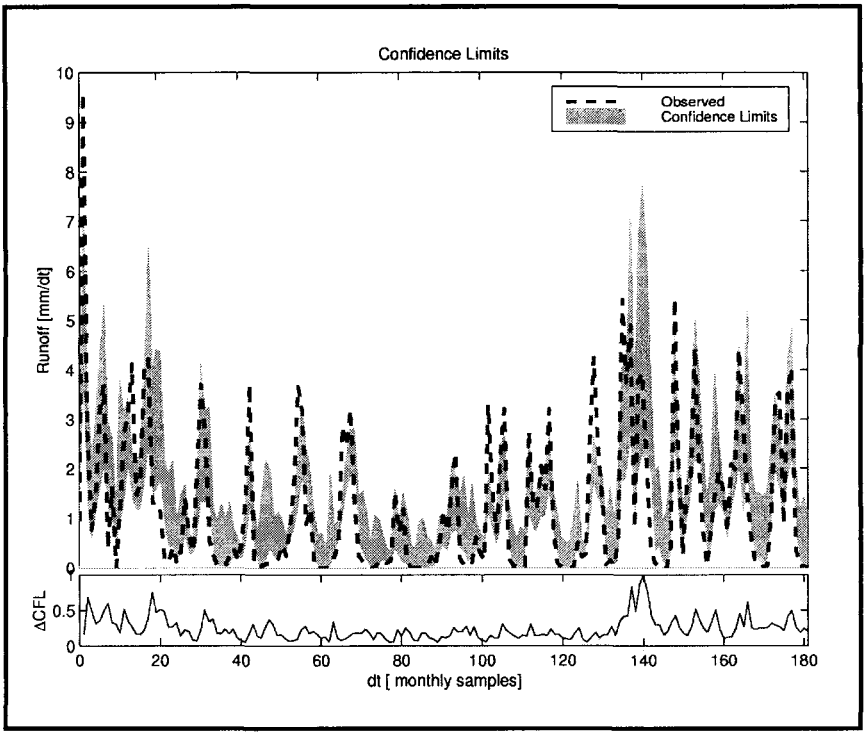


Fig. A.4 Plot showing 95% confidence limits, calculated in the Box-Cox transformed space. The graph is an example taken from the regionalization study.

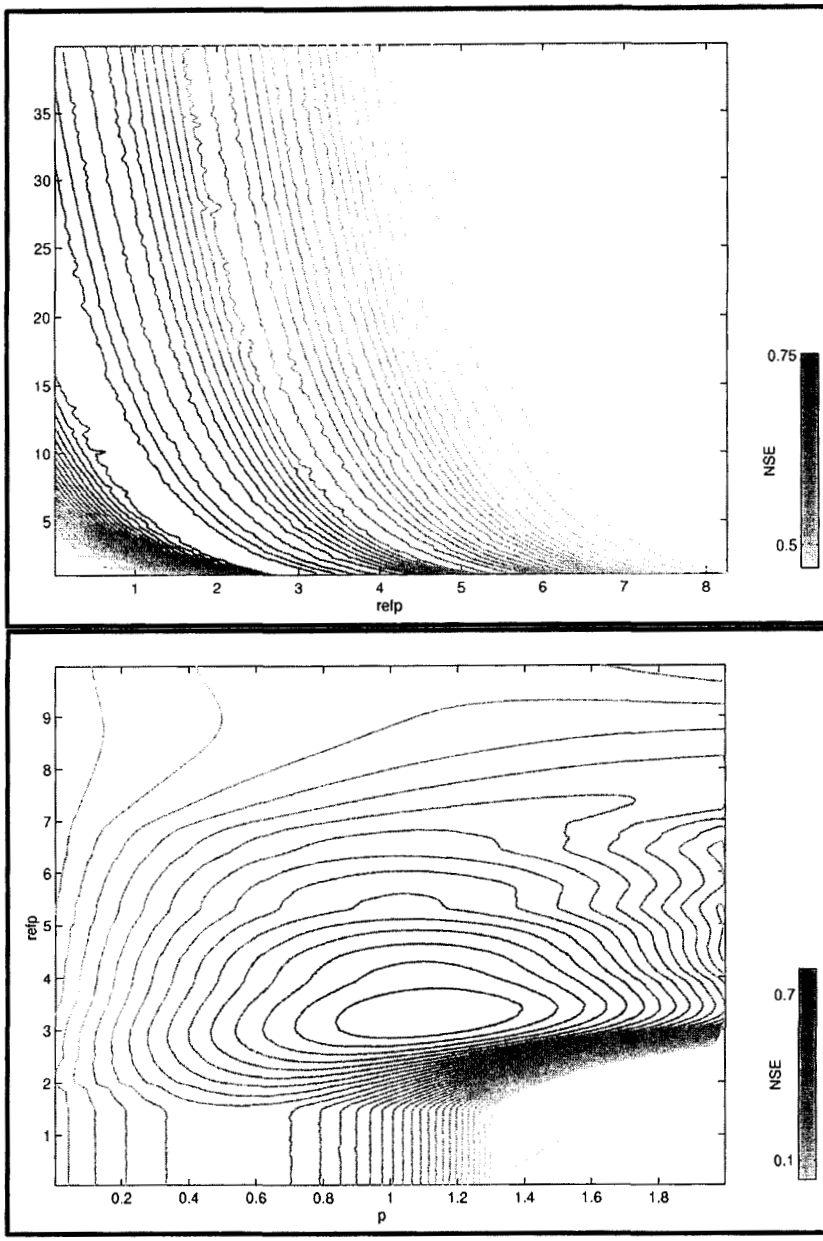


Fig. A.5 Two-dimensional MCAT plots show the interaction between two parameters. The top plot shows a poorly identified parameter that was subsequently removed.

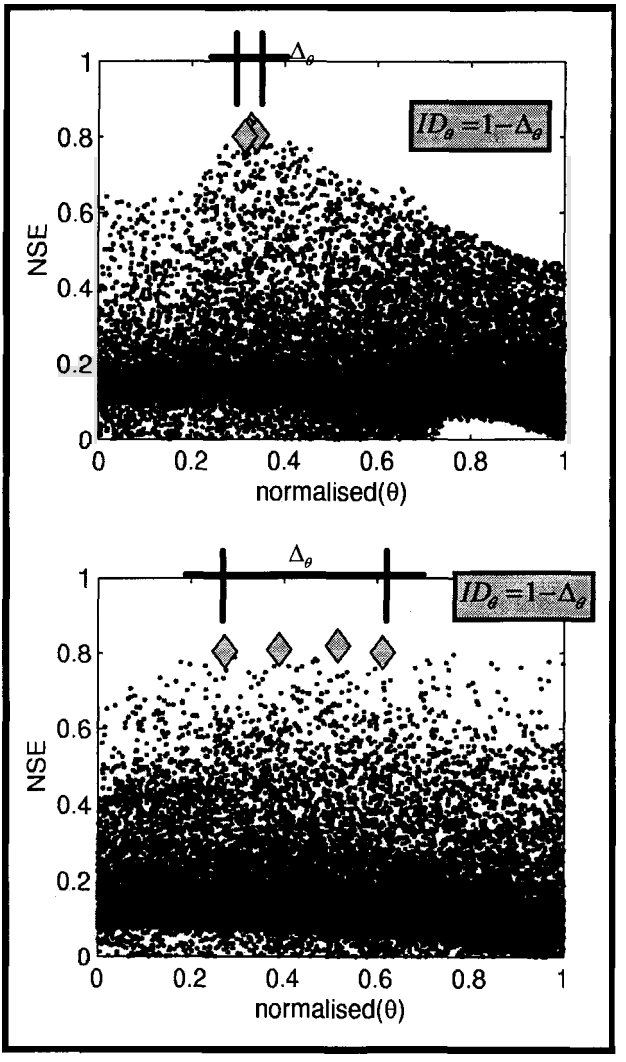


Fig. A.6 Calculation of the identifiability measure ID

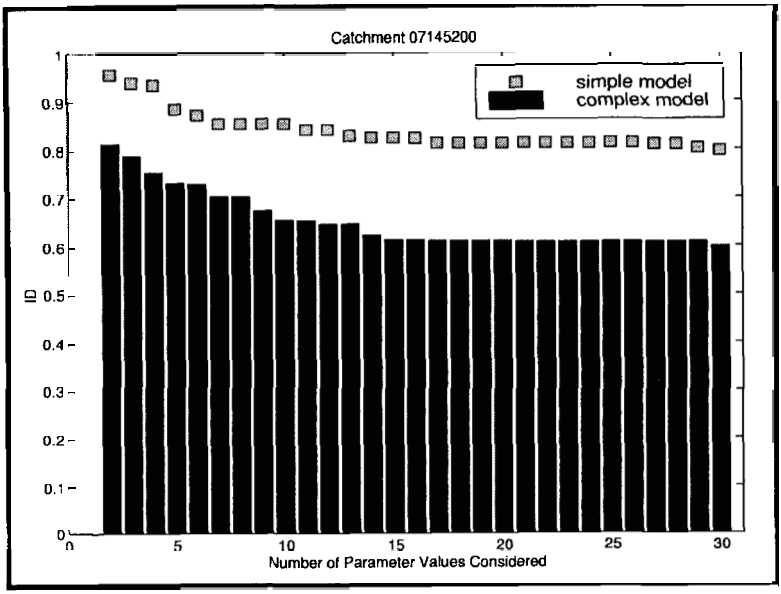


Fig. A.7 Identifiability comparison between complex and simple model for one example catchment.

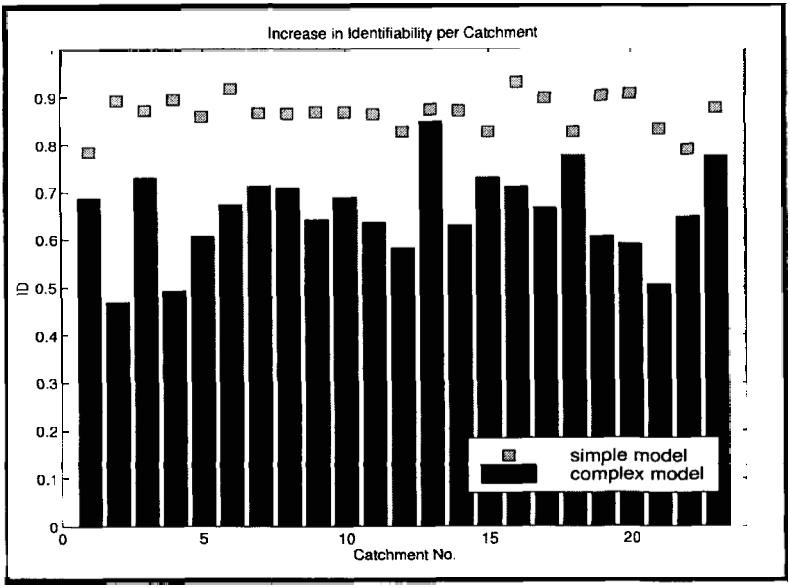


Fig. A.8 Increase in ID for all 23 catchments using the top 5 values.

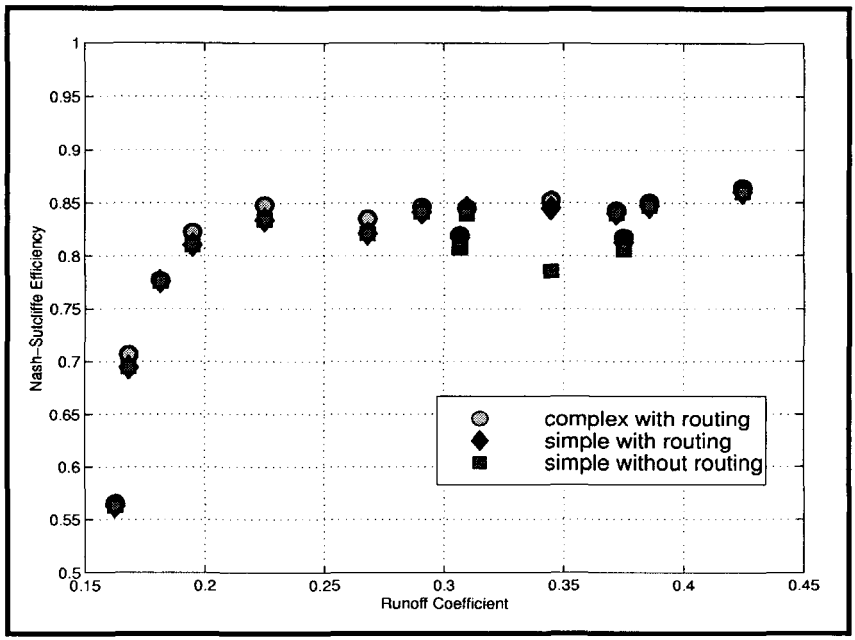


Fig. A.9 Model comparison results.

Parameter Sensitivity

Which components of a model structure (Fig. A.10) are active during a simulation changes depending on the dominant response mode of the hydrological system. Segmenting the observed system response with respect to the different modes (Fig. A.11) allows the modeller to formulate different objective functions for every mode. The most sensitive parameters for each mode can be visualized using the extended Regional Sensitivity Analysis described earlier (Fig. A.12).

This additional information can be used in a multi-objective optimization framework to increase the amount of information that can be retrieved from the available data. These are preliminary results and a more detailed investigation is required to find the best way of using this additional information in the calibration process.

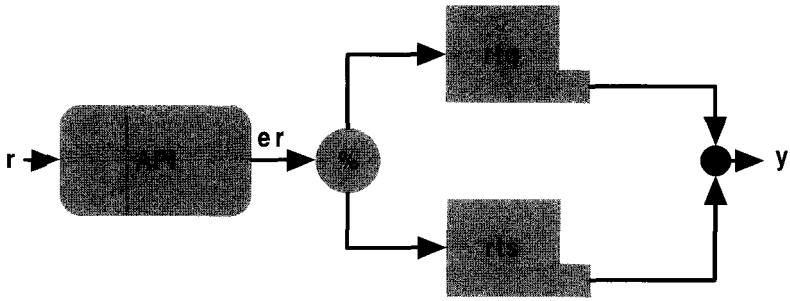


Fig. A.10 Model structure used: API and two linear reservoirs.

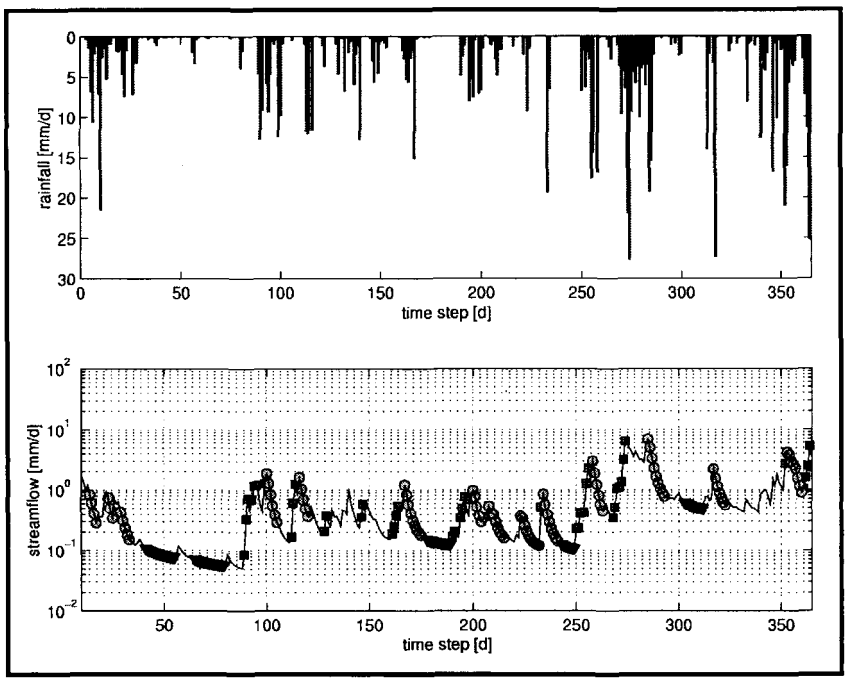


Fig. A.11 Measured rainfall and segmented synthetic flow used.

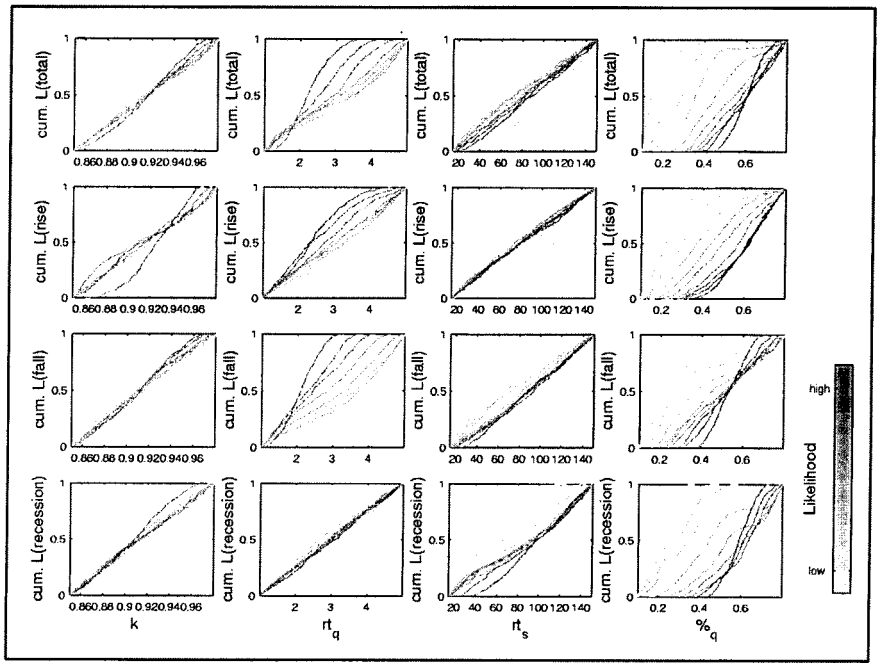


Fig. A.12 Visualization of changing sensitivity using extended RSA plots. The parameters are the API decay parameter k, the two residence times (rt_q and rt_s) and the percentage quick flow (%q).

Appendix B

The Rainfall-Runoff Modelling Toolbox

Introduction

The rainfall-runoff modelling toolbox (RRMT) was used for all simulations throughout this monograph and can be downloaded free of charge for non-commercial use, *i.e.* research and teaching. A basic knowledge of Matlab is assumed. Books introducing Matlab can be found on http://www.mathworks.com.

The Rainfall-Runoff Modelling Toolbox – RRMT

This toolbox simulates the rainfall-runoff relationship in natural catchments. The objective of the RRMT is to develop and test parsimonious and lumped, parametric or hybrid metric-parametric model structures with a high level of parameter identifiability. It is a modelling framework with a modular structure. It also offers tools for data analysis and manipulation, parameter uncertainty analysis, plots for residual analysis etc.

Hard- and software requirements

Hardware

- Any computer capable of running Matlab version 5.2 or higher (Mathworks, 1996).

Software

- Any platform supporting Matlab 5.2 or higher. The RRMT was developed within a Windows NT environment.
- Matlab version 5.2 or higher. The RRMT was developed using Matlab 5.2. It might therefore not work properly in older versions.
- The toolbox is provided as a collection of P-files produced for different versions of Matlab.

Download

The RRMT and a corresponding user manual can be downloaded from the web site of the Environmental and Water Resource Engineering (EWRE) section of the Department of Civil and Environmental Engineering at the Imperial College London (UK) at http://ewre.cv.ic.ac.uk/software.

Appendix C

The Monte Carlo Analysis Toolbox

Introduction

The Monte Carlo analysis toolbox (MCAT) was used for analysis throughout this monograph and can be downloaded free of charge for non-commercial use, *i.e.* research and teaching. A basic knowledge of Matlab is assumed. Books introducing Matlab can be found on http://www.mathworks.com.

The Monte Carlo Analysis Toolbox – MCAT

This toolbox is a collection of analysis and visualization functions integrated through a graphical user interface. It can be used to analyse the results of Monte Carlo parameter sampling experiments or from population evolution approaches. A number of powerful techniques are included to investigate the structure, sensitivity, and parameter and output uncertainty of mathematical models. Although it has been developed within the context of ongoing hydrological research, all functions can be used to investigate any (dynamic) mathematical model.

Hard- and software requirements

Hardware

- Any computer capable of running Matlab version 5.2 or higher (Mathworks, 1996).

Software

- Any platform supporting Matlab 5.2 or higher. The MCAT was developed within a Windows NT environment.
- Matlab version 5.2 or higher. The MCAT was developed using Matlab 5.2, and may not work properly in older versions.
- The toolbox is provided as a collection of P-files produced for different versions of Matlab.

Download

The MCAT and a corresponding user manual can be downloaded from the web site of the Environmental and Water Resource Engineering (EWRE) section of the Department of Civil and Environmental Engineering at the Imperial College London (UK) at http://ewre.cv.ic.ac.uk/software.

Notation

Lowercase Roman symbols

%(q)	percentage quickflow contribution
1l	1 linear reservoir
2pll	2 parallel linear reservoirs
2pln	2 parallel reservoirs, one linear and the other non-linear
a	storage coefficient
ae	actual evapotranspiration
ae_l	actual evapotranspiration from lower store
ae_u	actual evapotranspiration from upper store
a_i, b_i	TF parameters
alpha	fraction of effective rainfall contributing to quick response
b, b	degree of spatial variability of storage capacity (shape parameter)
b_e	evapotranspiration parameter
buc	conceptual bucket element
bypass	bypass parameter
c	actual storage capacity
c	catchment
$c1$	evapotranspiration parameter
$c2$	evapotranspiration parameter
$c3$	maximum drainage that can occur whilst a moisture deficit exists
$c4$	maximum cmd that can occur before water ceases draining to the stream
cm1	SMA module based on cmd structure
cmax	maximum storage capacity
cmd	catchment moisture deficit
d	drainage
$dist_{ij}$	Euclidean distance measure
f	activation function in ANN
$f(c)$	probability density function of storage capacity
g	Penman fraction of PE
HBV	Hydrologiska Byråns Vattenblanasavdelining model structure
i	index
ic1	SMA module based on Penman (pen) structure
j	index
k	time step index
k	API loss parameter
k(quick)	quickflow residence time
k(slow)	slowflow residence time
k_b	groundwater recession constant
k_i	reservoir time constant
m	TF index
md_1	moisture deficit in upper store
md_2	moisture deficit in lower store
mf	temperature or PE modulation factor
ms	catchment wetness index
n	vector length
nl	coefficient of non-linearity
p	number of parameters

pdx	probability distributed stores SMA module, x is an index
pen	SMA module based on Penman model structure
q	reservoir outflow
r	autocorrelation coefficient
r	rainfall
*r**	net-rainfall
rc	root constant
rd	rank difference
refp	reference parameter
r_s	Spearman rank correlation coefficient
rt(q)	quickflow residence time
rt(s)	slowflow residence time
s	actual (soil moisture) storage content
s_{max}	maximum storage content
s_{max1}	size of upper store
s_{max2}	size of lower store
t	time step
u	effective rainfall
v	volumetric constant
w	weight
x	filtered input to ANN
$\hat{y}$	calculated flow
y	observed flow
z^{-1}	backward shift operator

Uppercase Roman symbols

A [AREA]	total catchment area
A( ), B( )	TF polynomials
A_c	contributing area
ACTE	local model parameter describing the actual evapotranspiration
AET	rate of actual evapotranspiration period
ALPHA	percentage quickflow contribution
AREA	catchment size
ASPBAR	index representing the dominant aspect of catchment slopes
ASPVAR	index describing the invariability in aspect of catchment slopes
B	behavioural
$\overline{B}$	non-behavioural
BEXP	shape parameter
BFIHOST	base flow index from HOST
BIAS	bias
C	square root of Σ^{-1}
C	rational formula coefficient
CMAX	maximum storage capacity
DPLBAR	index describing catchment size and drainage path configuration
DV	deviation of runoff volumes
ER	effective rainfall
F( )	cumulative distribution function
F(c)	distribution function of storage capacity
FARL	index of flood attenuation due to reservoirs and lakes
FC	field capacity
FD	RMSE for driven flow period
FD, FQ, FS	RMSE for response modes
FDH	RMSE for flow periods that are driven and high
FDL	RMSE for flow periods that are driven and low
FH, FM, FL	RMSE for horizontal segmentation
FQ	RMSE for non-driven quick-flow period
FS	RMSE for non-driven slow-flow rainfall

FWU *RMSE* for warming up period
G gradient distribution
HMLE heteroscedastic maximum likelihood estimator
I meteorological forcing variables (e.g. rainfall)
ID identifiability measure
Kq quickflow reservoir coefficient
Ks slowflow reservoir coefficient
LDP longest drainage path
LR linear reservoir
M model structure
M_L local model structure
M_R regional model structure
N vector length
N, *K* NIUH parameters
NSE Nash-Sutcliffe efficiency
P, *P* precipitation
PE potential evapotranspiration
PEF measure of performance
Pi parameter, with i being an index
PROPWET index of proportion of time that soils are wet
Q streamflow
Q_{total} rate of surface and subsurface runoff
Q_D driven flow period
Q_Q non-driven quick-flow period
Q_S non-driven slow-flow period
QUICK local model parameter describing
t he quick response residence time
R^2 coefficient of determination
RMED-1D median annual maximum 1-day rainfall
RMED-1H median annual maximum 1-hour
RMED-2D median annual maximum 2-day rainfall
RMSE root mean squared error
S storage
SAAR 1961-90 standard-period average annual rainfall
$SAAR_{4170}$ 1941-70 standard-period average annual rainfall
SLOW local model parameter describing the slow response residence time
SLS simple least squares
SPLIT local model parameter describing the split of the effective rainfall into quick and slow response
SPRHOST standard percentage runoff from HOST
SS_E residual sum of squares
STORE local model parameter describing the storage volume
T residence time
TF transfer function
URBLOC index of location of urban and suburban land cover
$URBTEXT_{1990}$ FEH index of fractional urban extend for 1990
URS uniform random search
V volume of water stored in catchment
W transformed catchment characteristics matrix in weighted regression
WP wilting point
Z transformed dependent variable vector in weighted regression

Lowercase Greek symbols

α	percentage quickflow contribution
α_{CFL}	percentile
$\hat{\sigma}^2$	error variance
ε_R	error (residual) term for regional model
ε_L	error (residual) term for local model
θ_L	local parameter (vector)
$\hat{\theta}_L$	local parameter estimated by regional model
θ_R	regional (regression) parameter (vector)
μ	estimated mean
λ	Box-Cox transformation parameter
ϕ	vector of physiographical and meteorological catchment characteristics
σ	standard deviation
δ	transformed error term in weighted regression
τ	depletion rate through losses to stream and evapotranspiration
τ_w	time constant of catchment losses at *RT*
ρ	bypass parameter
γ	groundwater recharge
δ	TF delay
ε	residual
θ	model parameter set or vector
λ	(Box-Cox) transformation parameter
τ	lag
ς	support

Uppercase Greek symbols

Λ	weight matrix in weighted regression
Φ	matrix of physiographical and meteorological catchment characteristics

Acronyms

ANN	artificial neural network
API	antecedent precipitation index
ARS	adaptive random search
CRR	conceptual rainfall-runoff
CRS	controlled random search
DYNIA	dynamic identifiability analysis
EKF	extended Kalman filter
FEH	Flood Estimation Handbook
FR	fuzzy regression
GA	genetic algorithm
GLUE	generalised likelihood uncertainty estimation
GP	genetic programming
HOST	hydrology of soil types classification
IHDTM	Institute of Hydrology digital terrain model
LHS	latin hypercube sampling
MCAT	Monte Carlo analysis toolbox
MCMC	Markov chain Monte Carlo
MCSM	Monte Carlo set membership
ML	maximum likelihood
MO	multi-objective
MOCOM	multi-objective complex evolution algorithm
MVR	multivariate regression
NERC	natural environment research council

NIUH	Nash Instantaneous Unit Hydrograph
NWS	National Weather Service
OF	objective function
PU	prediction uncertainty
RR	rainfall-runoff
RRMT	rainfall-runoff modelling toolbox
RSA	regional sensitivity analysis
SCE	shuffled complex evolution algorithm
SCS-CN	soil conservation service curve number
SMA	soil moisture accounting
UH	unit hydrograph
URS	uniform random search
WMO	World Meteorological Organisation
WY	water-years

This page intentionally left blank

Bibliography

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. and Rasmussen, J. (1986). An introduction to the European Hydrological System – Système Hydrologique Europeén, SHE. 1. History and philosophy of a physically-based, distributed modelling system. *J. Hydrol.*, 87, pp. 45-59.

Abdulla, F. A. and Lettenmaier, D. P. (1997). Development of regional parameter estimation equations for a macroscale hydrologic model, *J. Hydrol.*, 197, pp. 230-257.

Abdulla, F. A., Lettenmaier, D. P. and Liang, X. (1999). Estimation of the ARNO model baseflow parameters using daily streamflow data, *J. Hydrol.*, 222, pp. 37-54.

Acreman, M. C. and Sinclair, C. D. (1986). Classification of drainage basins according to their physical characteristics: an application for flood frequency analysis in Scotland. *J. Hydrol.*, 217, pp. 239-252.

Akaike, H. (1974). A new look at statistical model identification. *IEEE Trans. Auto. Control*, AC-19, pp. 716-723.

Alley, W. M. (1984). On the treatment of evapotranspiration, soil moisture accounting, and aquifer recharge in monthly water balance models, *Water Resour. Res.*, 20(8), pp. 1137-1149.

Ambroise, B., Beven, K. J. and Freer, J. (1996). Toward a generalization of the TOPMODEL concepts: Topographic indices of hydrological similarity, *Water Resour. Res.*, 32(7), pp. 2135-2145.

Anderson, M. G. and Bates, P. D., eds. (2001). *Validation in hydrological modelling*, John Wiley & Sons, New York, pp. 500.

Anderson, M.G. and Burt, T.P. (1985). Modelling strategies. In Anderson, M. G. and Burt, T.P. (eds.), *Hydrological forecasting*, John Wiley & Sons, New York, pp. 1-13.

Anderson, M.P. and Woesner, W.W. 1992. *Applied groundwater modelling. Simulation of flow and advective transport.* Academy Press, San Diegeo, USA.

Ando, Y. (1990). Regionalisation of parameters using basin geology, land use, and soil type for use in a storm rainfall-runoff relationship. In *Regionalisation in Hydrology*, IAHS Publ. 191, pp. 211-218.

Arch, S. E. (1997). The role of spatially varying transmissivity on groundwater and streamflow response in a chalk aquifer. Unpublished M.Sc. Thesis, Department of

Civil Engineering, Imperial College of Science, Technology and Medicine, London, UK.

ASCE (1993). Criteria for evaluation of watershed models. *J. Irrigation and Drainage Eng.*, 119(3), 429-442.

Atkinson, S. E. (2001). Climate, soil vegetation controls on streamflow variability with changing timescales and implications for model complexity. Unpublished M.Sc. Thesis, Department of Environmental Engineering, University of Western Australia, Nedlands, Australia.

Babovic, V. and Bojkov, V. H. (2001). Runoff modelling with genetic programming and artificial neural networks, D2K Unpublished Technical Report, D2K TR 0401-1, Danish Hydraulic Institute, DK.

Bard, Y. (1974) *Non-linear parameter estimation*, Academic Press, New York, N.Y.

Bardossy, A. (1990). Note on fuzzy regression. *Fuzzy Sets and Systems*, 37(1), 65-75.

Bardossy, A., Bogardi, I. and Duckstein, L. (1990). Fuzzy regression in hydrology. *Water Resour. Res.*, 26(7), pp. 1497-1508.

Bashford, K. E., Beven, K. J. and Young, P. C. (2002). Observational data and scale-dependent parameterizations: explorations using a virtual hydrological reality. *Hydrol. Processes*, 16, pp. 293-312.

Bastidas, L. A. (1998). Parameter estimation for hydrometeorological models using multi-criteria methods. Unpublished Ph.D. Thesis, University of Arizona, Tucson, Ariz., U.S.A.

Bastidas, L. A., Gupta, H. V., Sorooshian, S., Shuttleworth, W. J. and Yang, Z .L. (1999). Sensitivity analysis of a land surface scheme using multicriteria methods. *J. of Geophys. Res.*, 104(D16), 19481-19490.

Bates, B. C. and Campbell, E. P. (2001). A Markov chain Monte-Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modelling. *Water Resour. Res.*, 37(4), pp. 937-947.

Bayliss, A. (1999). *Flood estimation handbook, 5: Catchment descriptors.* Institute of Hydrology, Wallingford, U.K.

Beck, M. B. (1981a). A procedure for modeling. In Orlob, G.T. (ed.) *Mathematical modeling of water quality: Streams, lakes, and reservoirs.* John Wiley & Sons, 11-41.

Beck, M. B. (1981b). Sensitivity analysis, calibration, and validation. In Orlob, G.T. (ed.), *Mathematical modeling of water quality: Streams, lakes, and reservoirs.* John Wiley & Sons, New York, N.Y., pp. 425-467.

Beck, M. B. (1985). Structures, failure, inference and prediction. In Barker, M.A. and Young, P.C. (eds.) *Identification and System Parameter Estimation: Proceedings 7th Symposium* Volume 2, July 1985, IFAC/IFORS, York, UK, pp. 1443-1448.

Beck, M. B. (1987). Water quality modelling: A review of the analysis of uncertainty. *Water Resour. Res.*, 23(8), pp. 1393-1442.

Beck, M. B., Jakeman, A. J. and McAleer, M. J. (1993). Construction and evaluation of models of environmental systems. In Jakeman, A. J., Beck, M. B. and McAleer,

M.J. (eds.), *Modelling change in environmental systems,* John Wiley and Sons, New York, N.Y., pp. 3-35.

Becker, A. (1992). Criteria for a hydrologically sound structuring of large scale land surface process models. In O'Kane, J.P. (ed.). *Advances in theoretical hydrology.* Elsevier, Amsterdam, NL, pp. 97-111.

Becker, A. and Nemec, J. (1987). Macroscale hydrologic models in support to climate research. In *The influence of climate change and climatic variability on the hydrologic regime and water resources.* Symposium Proceedings, Vancouver, CA, IAHS Publ. no. 168, pp. 431-445.

Bentura, P. L. F. and Michel, C. (1997). Flood routing in a wide channel with a quadratic lag-and-route method. *Hydrological Sciences Journal,* 42(2), pp. 169-189.

Bergström, S. (1995). The HBV Model. In Singh, V.P. (ed.) *Computer models of watershed hydrology.* Water Resources Publications, Colorado, USA, pp. 443-476.

Beven, K. J. (1985). Distributed models. In Anderson, M.G. and Burt, T.P. (eds.) *Hydrological forecasting.* John Wiley & Sons, Chichester, UK, pp. 405-435.

Beven, K. J. (1989). Changing ideas in hydrology – The case of physically-based models. *Journal of Hydrology,* 105, pp. 157-172.

Beven, K. J. (1998). Generalised Likelihood Uncertainty Estimation (GLUE) package v. 1.0 – User manual. University of Lancaster, UK.

Beven, K. J. (2000a). *Rainfall-runoff modeling: The primer.* John Wiley and Sons, Chichester, U.K, pp. 360.

Beven, K. J. (2000b). Uniqueness of place and process representations in hydrological modelling. *Hydrol. and Earth System Sci.,* 4(2), pp. 203-213.

Beven, K. J. (2001a). Calibration, validation and equifinality in hydrological modelling. Anderson, M.G. and Bates, P.D. (eds.), *Validation in hydrological modelling,* John Wiley & Sons, Chichester, U.K., pp. 43-55.

Beven, K. J. (2001b). How far can we go in distributed modelling? *Hydrology and Earth System Sciences,* 5(1), pp. 1-12.

Beven, K. J. (2002). Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes,* 16, pp. 189-206.

Beven, K. J. and Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin,* 24(1), pp. 43-69.

Beven, K. J. and Wood, E. F. (1983). Catchment geomorphology and the dynamics of runoff contributing areas. *J. Hydrol.,* 65, pp. 139-158.

Beven, K. J. and Binley, A. M. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes,* 6, pp. 279-298.

Beven, K. J. and Young, P. C. (2003). Comment on "Bayesian recursive parameter estimation for hydrologic models," by Thiemann, M., Trosset, M., Gupta, H. V. and Sorooshian, S. *Water Resour. Res,* 39(5).

Beven, K. J., Wood, E. F. and Sivapalan, M. (1988). On hydrological heterogeneity: Catchment morphology and catchment response. *J. Hydrol.,* 100, pp. 353-375.

Beven, K. J., Lamb, R., Quinn, P., Romanowicz, R. and Freer, J. (1995). Topmodel. In Singh, V. P. (ed.). *Computer models of watershed hydrology*. Water Resources Publications, Highlands Ranch, Colorado, USA, pp. 627-668.

Beven, K. J., Freer, J., Hankin, B. and Schulz, K. (2000). The use of generalised likelihood measures for uncertainty estimation in high-order models of environmental systems. In Fitzgerald, Smith, R. C., Walden, A. T. and Young, P. C. (eds.) *Nonlinear and nonstationary signal processing*. Cambridge University Press, UK.

Beveridge, W.I.B. (1957). *The art of scientific investigation*. 3rd edition, William Heinemann Ltd., Melbourne, Australia.

Binley, A. M., Beven, K. J. and Elgy, J. (1989). A physically-based model of heterogeneous hillslopes. II. Effective hydraulic conductivities. *Water Resour. Res.*, 25(6), pp. 1227-1233.

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 Resour. Res.*, 27(6), pp. 1253-1261.

Blackie, J. R. and Eeles, C. W. O. (1985). Lumped catchment models. In Anderson, M.G. and Burt, T.P. (eds.) *Hydrological forecasting*. John Wiley & Sons Ltd., pp. 311-345.

Boogaard, H. F. P. van den, Ali, M. S. and Mynett, A. E. (1998). Self organising feature maps for the analysis of hydrological and ecological data sets. In Babovic V. and Larsen L.C. (eds.) *Hydroinformatics'98*. Balkema, Rotterdam, NL, pp. 733-740.

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.

Box, G.E.P. and Cox, O.R. (1964). The analysis of transformation. *J. Royal Stat. Soc.*, Series B, 26(2), 211-252.

Box, G. P. and Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. Holden-Day, San Francisco, USA.

Boxall, S. (2001). Regionalisation of rainfall-runoff models for the Thames catchments. Unpublished M.Sc. Thesis, Imperial College of Science, Technology and Medicine, London, UK.

Boyle, D. P., Gupta, H. V. and Sorooshian, S. (2000). Toward improved calibration of hydrological models: Combining the strengths of manual and automatic methods. *Water Resour. Res.*, 36(12), 3663-3674.

Boyle D. P., Gupta H. V., Sorooshian S., Koren V., Zhang Z. and Smith M. (2001). Towards improved streamflow forecasts: The value of semi-distributed modelling. *Water Resour. Res.*, 37(11), 2739-2759.

Brazil, L.E. (1988). Multilevel calibration strategy for complex hydrologic simulation models. Unpublished Ph.D. Thesis, Colorado State University, Fort Collins, Colorado, USA.

Brazil, L.E. and Hudlow, M.D. (1981). Calibration procedures used with the National Weather Service Forecast System. In Hames, Y.Y. and Kindler, J. (eds.) *Water and related land resources systems*. Pergamon, New York, pp. 457-466.

Brazil, L. E. and Krajewski, W. F. (1987). Optimisation of complex hydrologic models using random search methods. Engineering Hydrology Proceedings, Hydraulics Division/ASCE/Williamsburg, Virginia, USA, 726-731.

Brun, R., Reichert, P. and Kuensch, H. R. (2001). Practical identifiability analysis of large environmental simulation models. *Water Resour. Res.*, 37(4), 1015-1030.

Burn, D. H. (1990). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resour. Res.*, 26(10), 2257-2265.

Burn, D H. and Boorman, D. B. (1992). Estimation of hydrological parameters at ungauged catchments. *J. Hydrol.*, 143, 429-454.

Burn, D. H. and Goel, N. K. (2000). The formation of groups for regional flood frequency analysis. *Hydrological Sciences Journal*, 45(1), 97-112.

Burnash, R. J. C. (1995). The NWS river forecast system – catchment modelling. In Singh, V. P. (ed.) *Computer models of watershed hydrology*. Water Resources Publications, Colorado, USA, 311-366.

Burnash, R. J. C., Ferral, R. L. and McGuire, R. A. (1973). A generalized streamflow simulation system: conceptual modeling for digital computers, U.S. Department of Commerce, National Weather Service and State of California, Department of Water Resources, Sacramento, California.

Calder, I. R., Harding, R. J. and Rosier, P. T. W. (1983). An objective assessment of soil-moisture deficit models. *J. Hydrol.*, 60, pp. 329-355.

Calver, A. (1988). Calibration, sensitivity and validation of a physically based rainfall-runoff model. *J. Hydrol.*, 103, pp. 102-115.

Calver, A. (1993). The time-area formulation revisited. Proceedings of the Institution of Civil Engineers for Water, Maritime & Energy, 101, pp. 31-36.

Calver, A. (1996). Development and experience of the 'TATE' rainfall-runoff model. *Proceedings of the Institution of Civil Engineers for Water, Maritime & Energy*, 118, 168-176.

Calver, A., Lamb, R. and Morris, S .E. (1999). River flood frequency estimation using continuous runoff modelling. *Proceedings of the Institution of Civil Engineers for Water, Maritime & Energy*, 136, pp. 225-234.

Camacho, L. A. (1997). Extension of the QUASAR Water Quality Model to Incorporate Dead Zone Mixing Concepts. Unpublished M.Sc. Thesis, Civil and Environmental Engineering Department, Imperial College of Science, Technology and Medicine, London, UK.

Cameron, D. (2000). Estimating flood frequency by continuous simulation. Unpublished Ph.D. Thesis, University of Lancaster, UK.

Cameron, D., Beven, K. J. and Naden, P. (2001). Flood frequency estimation under climate change (with uncertainty). *Hydrology and Earth System Sciences*, 4(3), pp. 393-405.

Campbell, E. P. and Bates, B.C. (2001). Regionalisation of rainfall-runoff parameters using Markov Chain Monte Carlo samples. *Water Resour. Res.*, 37(3), pp. 731-739.

Campbell, E.P., Fox, D.R. and Bates, B.C. (1999). A Bayesian approach to parameter estimation and pooling in nonlinear flood event models. *Water Resour. Res.*, 35(1), pp. 211-220.

Chang, M. and Boyer, D.C. (1977). Estimates of low flows using watershed and climatic parameters. *Water Resour. Res.*, 13(6), pp. 997-1001.

Chankong, V. and Haimes, Y.Y. (1993). Multi-objective optimisation: Pareto optimality. In Young, P.C. (ed.) *Concise encyclopaedia of environmental systems.* Pergamon Press, UK, pp. 387-396.

Chapman, T. (1999). Comparison of algorithms for stream flow recession and baseflow separation. *Hydrological Processes*, 13, 701-714.

Chow, V.T. and Kulandaiswamy, V.C. (1971). General hydrologic system model. *Journal of the Hydraulic Division*, American Society of Civil Engineers, 97(HY6), pp. 791-804.

Chow, V.T., Maidment, D.R. and Mays, L.W. (1988). *Applied Hydrology.* McGraw-Hill, USA.

Clarke, R.T. (1973). Mathematical Models in Hydrology. Irrigation and Drainage Paper 19 , FAO, Rome, Italy.

Clarke R.T. (1994). *Statistical Modelling in Hydrology.* John Wiley & Sons Ltd., UK.

Crawford, N.H. and Linsley, R.K. (1966). Digital simulation in hydrology: Stanford watershed model IV. Department of Civil Engineering, Stanford University, Technical Report No. 39, California, USA.

Dawdy, D.R. and O'Donnell, T. (1965). Mathematical models of catchment behaviour. *J. of the Hydraulic Division*, ASCE, 91(HY4), pp. 123-137.

Dawson, C. and Wilby, R. (1998). An artificial neural network approach to rainfall-runoff modelling. *Hydrological Sci. J.*, 43, pp. 47-66.

Demuth, H. and Beale, M. (1997. *Neural network toolbox – User's guide version 3.0.* The Mathworks Inc., Nattick, USA.

Dennis, J.E.Jr., Gay, D.M. and Welsh, R.E. (1981). An adaptive nonlinear least squares algorithm: Association for computing machinery. *Trans. Math. Software*, 7, pp. 348-368.

Dingman, S.L. (1994). *Physical Hydrology.* Prentice Hall, Englewood Cliffs, New Jersey, USA.

Diskin, M. H. and Simon, E. (1977). A procedure for the selection of objective functions for hydrologic simulation models. *J. Hydrol.*, 34, 129-149.

Dixon, L. C. W. (1972). *Nonlinear Optimisation.* The English Universities Press Limited, London.

Dooge, J. C. I. (1968). The hydrologic cycle as a closed system. *Bull. International Association of Hydrological Sciences*, 13(1), 58-68.

Dooge, J. C. I. (1978). General report on model structure and classification. In Askew, A.J., Greco, F. and Kindler, J. (eds.) *Logistics and benefits of using mathematical models of hydrologic and water resource systems.* IIASA Proceedings Series, Pergamon Press, Oxford, pp. 1-21.

Drane, N. (2000). The development of a simple conceptual rainfall-runoff model for an ephemeral catchment in Southern England: Assessing the impacts of climate change scenarios on the hydrological system. M.Sc. Thesis, Civil and Environmental Engineering Department, Imperial College of Science, Technology and Medicine, London, UK.

Draper, N. R. and Smith, H. (1981). *Applied regression analysis.* John Wiley and Sons, New York, USA.

Drecourt, J.-P. (1999). Application of neural networks and genetic programming to rainfall runoff modelling. D2K Unpublished Technical Report, D2K-0699-1, Danish Hydraulic Institute.

Duan, Q., Sorooshian, S. and Gupta, V. K. (1992). Effective and efficient global optimisation for Conceptual Rainfall-Runoff Models. *Water Resour. Res.,* 28(4), pp. 1015-1031.

Duan, Q., Sorooshian, S. and Gupta, V. K. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *J. of Optimisation Theory and Applications*, 76(3), pp. 501-521.

Duan, Q., Sorooshian, S. and Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.,* 158, pp. 265-284.

Duan, Q., Schaake, J. and Koren, V. (2001). *A priori* estimation of land surface model parameters. In Lakshmi, V., Albertson, J. and Schaake, J. (eds.) *Land surface hydrology, meteorology, and climate: Observations and modelling.* American Geophysical Union, Water Science and Application, Vol. 3, pp. 77-94.

Duckstein, L., Bogardi, I. and Bardossy, A. (1992). Research on fuzzy sets in hydrology, water resources and related areas. *Fuzzy Sets and Systems*, 46(3), pp. 385-386.

Dümenil, L. and Todini, E. (1992). A rainfall-runoff scheme for use in the Hamburg climate model. In O'Kane, J.P. (ed.) *Advances in theoretical hydrology.* Elsevier, Amsterdam, pp. 129-157.

Dunne, S.M. (1999). Imposing constraints on parameter values of a conceptual hydrological model using baseflow response, *Hydrology and Earth System Sciences*, 3, pp. 271-284.

Dunne, T. (1978). Field studies of hillslope flow processes. In Kirkby, M.J. (ed.) *Hillslope hydrology.* John Wiley & Sons, Chichester, UK, pp. 227-293.

Dyer, B. G., Nathan, R. J., McMahon, T. A. and O'Neill, I. C. (1994). Development of regional prediction equations for the RORB runoff routing model. Report 94/1, Cooperative research centre for catchment hydrology.

Edjiatno, N.N. De O., Yang, X., Makhlouf, Z. and Michel, C. (1999). GR3J: A daily watershed model with three free parameters. *Hydrological Sciences Journal*, 44(2), pp. 263-277.

Ehrgott, M. (2000). *Multicriteria optimisation*. Springer Verlag, Berlin.

Etter, D. M. (1997). *Engineering problem solving with Matlab* - 2nd Edition. Prentice-Hall, USA.

Evans, J. P. and Jakeman, A. J. (1998). Development of a simple, catchment-scale, rainfall-evapotranspiration-runoff model. *Environmental Modelling and Software*, 13, pp. 385-393.

Fernandez, W., Vogel, R.M. and Sankarasubramanian (2000). Regional calibration of a watershed model. *Hydrological Sciences J.*, 45(5), pp. 689-707.

Fluegel, W.-A. (1995). Delineating hydrological response units by geographical information system analysis for regional hydrological modelling using PRMS/MMS in the drainage basin of the river Broel, Germany. *Hydrological Processes*, 9, pp. 423-436.

Forsythe, G. E., Malcolm, M. A. and Moler, C. B. (1976). *Computer methods for mathematical computations*. Prentice-Hall, USA

Franchini, M. and Pacciani, M. (1991). Comparative analysis of several conceptual rainfall-runoff models. *J. Hydrol.*, 122, pp. 161-219.

Franchini, M., Galeati, G. and Berra, S. (1998). Global optimisation techniques for the calibration of conceptual rainfall-runoff models. *Hydrol. Sci. J.*, 43, pp. 443-458.

Franks, S, Gineste, P, Beven, K. J., and Merot, P. (1998). On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process. *Water Resour. Res.*, 34, pp. 787-797.

Franks, S. and Beven, K. J. (1997). Estimation of evapotranspiration at the landscape scale: a fuzzy disaggregation approach. *Water Resour. Res.*, 33, 2929-2938.

Freer, J., Beven, K. J. and Ambroise, B. (1996). Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. *Water Resour. Res.*, 32(7), pp. 2161-2173.

Freeze R. A. and Harlan R. L. 1969. Blueprint for a physically-based digitally simulated hydrologic response model. *J. Hydrol.*, 9, pp. 237-258.

Funke, R., Schumann, A. H. and Schultz, G.A. (1999). Regionalisation of parameters in mesoscale hydrological models. In Diekkrueger, B., Kirby, M.J. and Schroeder, U. (eds.) *Regionalisation in hydrology*. Conference Proceedings, Braunschweig, D, IAHS Publ. no. 254, pp. 171-179.

Gan, T. Y. and Biftu, G. F. (1996). Automatic calibration of conceptual rainfall-runoff models: Optimisation algorithms, catchment conditions, and model structure. *Water Resour. Res.*, 32(12), pp. 3513-3524.

Gan, T. Y. and Burges, S. J. (1990). An assessment of a conceptual rainfall-runoff model's ability to represent the dynamics of small hypothetical catchments 2. Hydrologic responses for normal and extreme rainfall. *Water Resour. Res.*, 26(7), pp. 1605-1619.

Gan, T. Y., Dlamini, E.M. and Biftu, G. F. (1997). Effects of model complexity and structure, data quality, and objective functions on hydrologic modelling, *J. Hydrol.*, 192, pp. 81-103.

Garrick, M., Cunnane, C. and Nash, J.E. (1978). A criterion of efficiency for rainfall-runoff models. *J. Hydrol.*, 36, pp. 375-381.

Gershenfeld, N. (1999). *The nature of mathematical modelling*. Cambridge University Press, Cambridge, UK.

Goldberg, D.E. (1989). Genetic algorithms in search, optimisation, and machine learning. Addison-Wesley, Reading, Mass.

Grayson, R. B., Moore, I. D. and McMahon, T.A. (1992). Physically-based hydrologic modelling. 2. Is the concept realistic. *Water Resour. Res.*, 28, pp. 2659-2666.

Green, C. S. (1979). An improved subcatchment model for the river Dee. Unpublished Report, Institute of Hydrology, Wallingford, No. 58.

Greenfield, B. (1984). The Thames water catchment model. Unpublished Report, Technology and Development Division, Thames Water, UK.

Grindley, J. (1970). Estimation and mapping of evaporation. In: *Symposium on world water balance*. Proceedings Reading Symposium, IAHS, Publ. no. 92, Vol. 1, pp. 200-213.

Grosbois, E. de, Hooper, R.P. and Christophersen, N. (1988). A multisignal automatic calibration methodology for hydrochemical models: A case study of the Birkenes model. *Water Resour. Res.*, 26, pp. 1299-1307.

Gupta, H. V (2001). The devilish Dr. M or Some comments on the identification of hydrologic models. British Hydrological Society Circulations, 68, pp. 1-4.

Gupta, V. K. and Sorooshian, S. (1983). Uniqueness and observability of conceptual rainfall-runoff model parameters: The percolation process examined. *Water Resour. Res.*, 19(1), pp. 269-276.

Gupta, V. K. and Sorooshian, S. (1985a). The relationship between data and the precision of parameter estimates of hydrologic models. *J. Hydrol.*, 81, pp. 57-77.

Gupta, V. K. and Sorooshian, S. (1985b). The automatic calibration of conceptual catchment models using derivative-based optimisation algorithms. *Water Resour. Res.*, 21(4), pp. 473-485.

Gupta, H. V., Sorooshian, S. and Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, 34(4), pp. 751-763.

Gupta, H. V., Thiemann, M. Trosset, M. and Sorooshian, S. (2003). Reply to comment by K. Beven and P. Young on "Bayesian recursive parameter estimation for hydrologic models." *Water Resour. Res.*, 39(5).

Gupta, H. V., Bastidas, L., Sorooshian, S., Shuttleworth, W. J. and Young, Z. L. (1999). Parameter estimation of a land surface sch3eme using multi-criteria methods. GCIP II Special Issue, *J. of Geophysical Res. Atmospheres,* 104(D16), pp. 19491-19503.

Hall, A.J. and Schaake, J. C. (1996). Model Parameter Estimation Experiment (MOPEX). Unpublished Report.

Hall, M J. and Minns, A. W. (1999). The classification of hydrologically homogeneous regions. *Hydrological Sciences J.*, 44(5), pp. 693-704.

Hanselman, D. and Littlefield, B. (1998). *Mastering Matlab 5 - A comprehensive tutorial and reference.* Prentice Hall, USA.

Harlin, J. (1991). Development of a process oriented calibration scheme for the HBV hydrological model. *Nordic Hydrology*, 22, pp. 15-36.

Harr, M. E. (1989). Probabilistic estimates for multivariate analyses. *Applied Mathematical Modelling*, 13, pp. 313-328.

Harvey, J. W. and Wagner, B. J. (2000). Quantifying hydrologic interactions between streams and their subsurface hyporheic zones. In Jones, J.A. and Mulholland, P.J. (eds.) *Streams and ground waters.* Academic Press, San Diego, pp. 3-44.

Helsel, D. R. and Hirsch, R. M. (1992). *Statistical methods in water resources.* Elsevier, Amsterdam, NL.

Hendrickson, J. D., Sorooshian, S. and Brazil, L. (1988). Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models. *Water Resour. Res.,* 24(5), pp. 691-700.

Herdegeen, R. G. and Reich, B. M. (1974). Unit hydrographs for catchments of different sizes and dissimilar regions. *J. Hydrol.,* 22, pp. 143-153.

Herschy, R. W. (1995). *Streamflow measurement,* 2nd edition, E & FN Spon, London, UK.

Hogue, T. S., Sorooshian, S., Gupta, H. V., Holz, A. and Braatz, D. (2000). A multistep automatic calibration scheme for river forecasting models. *J. of Hydrometeorology*, 1, pp. 524-542.

Holder, R. L. (1985). *Multiple regression in hydrology.* Institute of Hydrology, Wallingford, UK.

Holland, J. H. (1975). *Adaption in natural and artificial systems.* University of Michigan Press, Ann Arbor, Michigan, USA.

Hooke, R. and Jeeves, T.A. (1961). Direct search solutions of numerical and statistical problems. *J. Assoc. Comput. Mach.*, 8(2), pp. 212-229.

Hornberger, G.M. and Spear, R.C. (1981). An approach to the preliminary analysis of environmental systems, *J. of Environmental Management*, 12, pp. 7-18.

Hornberger, G.M., Beven, K.J., Cosby, B.J. and Sappington, D.E. (1985). Shenandoah watershed study: Calibration of a topography-based, variable contributing area hydrological model to a small forested catchment. *Water Resour. Res.,* 21, 1841-1859.

Hornberger, G.M., Raffensperger, J.P., Wiberg, P.L. and Eshleman, K.N. (1998). *Elements of hydrology.* John Hopkins University Press, Baltimore, USA.

Hough, M., Palmer, S., Weir, A., Lee, M. and Barrie, I. (1997). *The meteorological office rainfall and evaporation calculation system: MORECS version 2.0 (1995).* The Meteorological Office, Bracknell, UK.

Houghton-Carr, H. (1999). Flood estimation handbook – 4: Restatement and application of the flood studies report rainfall-runoff method. Institute of Hydrology, Wallingford, UK.

Høybye, J.A. (1996). Uncertainty modelling in hydrological applications. Unpublished Ph.D. Thesis, Lund University of Technology, S.

Hsu, K.-L., Gupta, H. V. and Sorooshian, S. (1995). Artificial neural network modelling of the rainfall-runoff process. *Water Resour. Res.,* 31(10), pp. 2517-2530.

Hughes, D. A. and Murrell, H. C. (1986). Non-linear runoff routing – A comparison of solution methods. *J. Hydrol.*, 85, pp. 339-347.

Ibitt, R. P. (1970). Systematic parameter fitting for conceptual models of catchment hydrology. Unpublished Ph.D. Thesis, Imperial College of Science, Technology and Medicine, London, UK.

Ibitt, R. P. and O'Donnell, T. (1971a. Designing conceptual catchment models for automatic fitting methods. IAHS Publication no. 101.

Ibitt, R. P. and O'Donnell, T. (1971b. Fitting methods for conceptual catchment models. *Journal of the Hydraulics Division*, 9, pp. 1331-1342.

Isabel, D. and Villeneuve, J.-P. (1986). Importance of the convergence criterion in the automatic calibration of hydrological models. *Water Resour. Res.,* 22(10), pp. 1367-1370.

Jakeman, A. J. and Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resour. Res.,* 29(8), pp. 2637-2649.

Jakeman, A. J., Littlewood, I. G. and Whitehead, P. G. (1990). Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *J. Hydrol.,* 117, pp. 275-300.

Jakeman, A. J., Hornberger, G. M., Littlewood, I. G., Whitehead, P. G., Harvey, J. W. and Bencala, K. E. (1992). A systematic approach to modelling the dynamic linkage of climate, physical catchment descriptors and hydrological response components. *Mathematics and Computers in Simulation*, 33, pp. 359-366.

Jakeman, A. J., Chen, T. H., Post, D. A., Hornberger, G. M., Littlewood, I. G. and Whitehead, P. G. (1993). Assessing uncertainties in hydrological response to climate at large scale. Macroscale modelling of the hydrosphere. Proceedings of the Yokohama Symposium, IAHS Publication No. 214, pp. 37-47.

Jakeman, A. J., Post, D. A. and Beck, M. B. (1994). From data and theory to environmental model: the case of rainfall runoff. *Environmetrics*, 5, pp. 297-314.

James, L. D. 1973. Hydrologic modelling, parameter estimation, and watershed characteristics. *J. Hydrol.,* 17, 283-307.

Jarboe, J. E. and Haan, C. T. (1974. Calibrating a water yield model from small ungauged catchments. *Water Resour. Res.,* 10(2), pp. 256-262.

Joerin, C., Musy, A. and Pointet, F. (1999). Hydrological processes identification by the combination of environmental tracing with time domain reflectrometry measurements. In Leibundgut C., McDonnell, J. and Schutlz, G. (eds.) *Integrated methods*

in catchment hydrology – Tracer, remote sensing and new hydrometric techniques. IAHS Publication no. 258, pp. 111-117.

Johnston, P. R. and Pilgrim, D. H. (1976). Parameter optimisation for watershed models. *Water Resour. Res.*, 12(3), pp. 477-486.

Jolley, T. J. (1995). Large-scale hydrological modeling: The development and validation of improved land-surface parameterizations for meteorological input. Unpublished Ph.D. Thesis, Imperial College of Science, Technology and Medicine, London, UK.

Jolley, T. J. and Wheater, H. S. (1997). An investigation into the effect of spatial scale on the performance of a one-dimensional water balance model. *Hydrological Processes*, 11, pp. 1927-1944.

Karnopp, D. C. (1963). Random search techniques for optimisation problems. *Automatica*, 1, pp. 111-121.

Kaufmann, A. and Gupta, M. M. (1988). *Fuzzy mathematical models in engineering and management science*. North-Holland, Amsterdam, NL.

Keesman, K. (1989). On the dominance of parameters in structural models of ill-defined systems. *Applied Mathematics and Computation*, 30, pp. 133-147.

Keijzer, M. and Babovic, V. (2000a). Genetic programming and the bias/variance tradeoff. Proceedings of the European Genetic Programming Conference, Edinburgh, UK.

Keijzer, M. and Babovic, V. (2000b). Genetic programming in prediction and knowledge discovery. Presentation given at the Data Mining Workshop, Hydroinformatics Conference, Iowa, USA.

Khodatalab, N., Gupta, H.V., Wagener, T. and Sorooshian, S. 2003. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *J. Hydrol.*, in Review.

Khu, S.-T. (1998). Automatic Calibration of NAM Model with Multi-Objectives Consideration. National University of Singapore / Danish Hydraulic Institute, D2K Technical Report 1298-1, DK.

Kilsby, C. G. and O'Connell, P. E. (2001). Flood Risk Estimation. Unpublished Lecture Notes, University of Newcastle, Newcastle-upon-Tyne, UK.

Kirkby, M. (1975). Hydrograph modelling strategies. In Peel, R., Chisholm, M. and Haggett, P. (eds.) Processes in physical and human geography. Heinemann, London, 69-90.

Kirkby, M. J. (1985). Hillslope hydrology. In Anderson, M. G. and Burt, T. P. (eds.) *Hydrological forecasting*. John Wiley and Sons, Chichester, UK, pp. 37-75.

Kirkpatrick, S., Gelatt Jr., C. D. and Vecchi, M. P. (1983). Optimisation by Simulated Annealing. *Science*, 220, pp. 671-680.

Kitanidis, P. K. and Bras, R. L. (1980a). Real-time forecasting with a conceptual hydrologic model 1. Analysis of uncertainty. *Water Resour. Res.*, 16(6), pp. 1025-1033.

Kitanidis, P. K. and Bras, R. L. (1980b). Real-time forecasting with a conceptual hydrologic model 2. Applications and results. *Water Resour. Res.*, 16(6), pp. 1034-1044.

Kleissen, F. M. (1990). Uncertainty and identifiability in conceptual models of surface water acidification. Unpublished PhD. Thesis, Imperial College of Science, Technology and Medicine, London, UK.

Kleissen F. M., Beck M. B. and Wheater, H. S. (1990). Identifiability of conceptual hydrochemical models. *Water Resour. Res.*, 26(2), pp. 2979-2992.

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences J.*, 31(1), pp. 13-24.

Klepper, O., Scholten, H. and van de Kamer, J. P. G. (1991). Prediction uncertainty in an ecological model of the Oosterschelde Estuary. *J. Forecasting*, 10, pp. 191-209.

Kokkonen, T. S. and Jakeman, A. J. (2001). A comparison of metric and conceptual approaches in rainfall-runoff modelling and its implications. *Water Resour. Res.*, 37(9), pp. 2345-2352.

Koren, V. I., Smith, M., Wang, D. and Zhang, Z. (2000). Use of soil property data in the derivation of conceptual rainfall-runoff model parameters. 15th Conference on Hydrology, Long Beach, American Meteorological Society, Paper 2.16, USA.

Kottegoda, N. T. and Rosso, R. (1997). Statistics, probability, and reliability for civil and environmental engineers. McGraw-Hill, New York, USA.

Koza, J. R. (1992). Genetic programming: On the programming of computers by means of natural selection. MIT Press, Cambridge, MA, USA.

Koza, J. R. (2000). What is genetic programming? http://www.genetic-programming.com/ gpanimatedtutorial.html

Kuczera, G. (1982). Combining site-specific and regional information: An empirical Bayes approach. *Water Resour. Res.*, 18(2), pp. 306-314.

Kuczera, G. (1983). Improved parameter inference in catchment models. 1. Evaluating parameter uncertainty. *Water Resour. Res.*, 19(5), pp. 1151-1162.

Kuczera, G. (1997). Efficient subspace probabilistic parameter optimisation for catchment models. *Water Resour. Res.*, 33(1), pp. 177-185.

Kuczera, G. and Mroczkowski, M. (1998). Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. *Water Resour. Res.*, 34(6), pp. 1481-1489.

Kuczera, G. and Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *J. Hydrol.*, 211, pp. 69-85.

Lamb, R. (2000). An approach to the calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation. *Water Resour. Res.*, 35, pp. 3103-3114.

Lamb, R. and Beven, K. (1997). Using interactive recession curve analysis to specify a general catchment storage model. *Hydrology and Earth System Sciences*, 1, pp. 101-113.

Lamb, R. and Calver, A. (2002). Continuous simulation as a basis for national flood frequency estimation. In Littlewood, I. (eds.) *Continuous river flow simulation: meth-*

ods, applications and uncertainties. British Hydrological Society - Occasional Papers, No. 13, Wallingford, UK, pp. 67-75.

Lamb, R., Beven, K. J. and Myrabo, S. (1998). Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model. *Advances in Water Resources*, 22(4), pp. 305-317.

Lamb, R., Crewett, J. and Calver, A. (2000). Relating hydrological model parameters and catchment properties to estimate flood frequencies from simulated river flows. BHS 7th National Hydrology Symposium, Newcastle-upon-Tyne, pp. 3.57-3.64.

Laurenson, E. M. (1964). A Catchment Storage Model for Runoff Routing. *J. Hydrol.*, 2, pp. 141-163.

Laurenson, E. M. and Mein, R. G. (1995). RORB: Hydrograph synthesis by runoff routing. In Singh, V.P. (ed.) *Computer models of watershed hydrology.* Water Resources Publications, Colorado, USA, pp. 151-164.

Leavesley, G. H., Markstrom, S. L., Restrepo, P. J. and Viger, R. J. (2002). A modular approach to addressing model design, scale, and parameter estimation issues in distributed hydrological modelling. *Hydrological Processes*, 16, pp. 173-187.

Lee, H. (2001). Rainfall-runoff model regionalisation in south-east England including uncertainty. Unpublished MSc. Thesis, Imperial College of Science, Technology and Medicine, London, UK.

Lees, M. J. (2000). Data-based mechanistic modelling and forecasting of hydrological systems. *J. Hydroinformatics*, 02.1, pp. 15-34.

Lees, M. J. and Wagener, T. (2000a). Review of HYCOM rainfall-runoff modeling undertaken for the review of Southern Water's Source Yields, (1997). Unpublished Report to Southern Water, Imperial College Consultants, London, UK.

Lees, M. J. and Wagener, T. (2000b). HYCOM rainfall-runoff modelling of Bewl-Darwell reservoir source rivers. Unpublished Report to Southern Water, Imperial College Consultants, London, UK.

Lees M. J., Price N., Wheater H. S. and Peach D. (1998). A Rainfall-runoff simulation model for South Galway, Ireland. In Proceedings of the BHS International Symposium on *Hydrology in a Changing Environment*, Vol. III, John Wiley & Sons, Chichester, UK, pp. 93-104.

Legates, D. R. and McCabe Jr., G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.*, 35(1), pp. 233-241.

Liong, S.-Y., Khu, S.-T. and Chan, W.-T. (2001). Derivation of Pareto front with genetic algorithm and neural network. *Journal of Hydrologic Engineering*, 6(1), pp. 52-61.

Linsley, R. K., Jr., Kohler, M. A. and Paulhus, J. L. H. 1949. *Applied hydrology.* McGraw-Hill, New York, USA.

Littlewood, I. G. and Jakeman, A. J. (1992). Characterisation of quick and slow streamflow components by unit hydrographs for single- and multi-basin studies. In Robinson, M. (ed.). *Methods of hydrological basin comparison.* Institute of Hydrology, Report No. 120, Wallingford, UK, pp. 99-111.

League, K. (1990). R-5 revisited. 2. Reevaluation of a quasi-physically based rainfall-runoff model with supplemental information. *Water Resour. Res.*, 26, pp. 973-987.

League, K. M. and Freeze, R. A. (1985). A comparison of rainfall-runoff modeling techniques on small upland catchments, *Water Resour. Res.*, 21(2), pp. 229-248.

Loon, E. E. van and Keesman, K. J. (2000). Identifying scale-dependent models: The case of overland flow at the hillslope scale. *Water Resour. Res.*, 36(1), 243-254.

Luce, C. H. and Cundy, T. W. (1994). Parameter Identification for a Runoff Model for Forest Roads. *Water Resour. Res.*, 30(4), pp. 1057-1069.

Macready, W. G. and Wolpert, D. H. (1996). What makes an optimisation problem hard? *Complexity*, 1, pp. 40-46.

Madsen, H. (2000a). Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *J. Hydrol.*, 235(3-4), pp. 276-288.

Madsen, H. (2000b). Automatic calibration and uncertainty assessment in rainfall-runoff modelling. Proceedings Joint Conference on Water Resources Engineering and Water Resources Planning and Management, Minneapolis, USA.

Madsen, H. and Jacobsen, T. (2001). Automatic calibration of the MIKE SHE integrated hydrological modelling system. Unpublished paper, 4th DHI Software Conference, Helsingør, DK.

Magee, B. (1977). *Popper*. Fontana/Collins, Glasgow, Scotland.

Magette, W. L., Shanholz, V. O. and Carr, J. C. (1976). Estimating selected parameters for the Kentucky watershed model from catchment characteristics. *Water Resour. Res.*, 12(3), pp. 472-476.

Mailhot, A., Gaume, E. and Villeneuve, J.-P. (1997). Uncertainty analysis of calibrated parameter values of an urban storm water quality model using Metropolis Monte Carlo algorithm. *Water Science and Technology*, 36(5), pp. 141-148.

Manabe, S. (1969). Climate and the ocean circulation 1. The atmosphere circulation and the hydrology of the earth's surface. *Monthly Weather Review*, 97(11), pp. 739-774.

Mankin, J. B., O'Neill, R. V., Shugart, H. H. and Rust, B. W. (1977). The importance of validation in ecosystem analysis. In Innis, G.S. (ed.) *New directions in the analysis of ecological systems*. Proceeding Series, 5(1), Simulation Council, La Jolla, California, USA, pp. 63-72.

Manley, R. E. (1978). Simulation of flows in ungauged basins. *Hydrological Sciences Bulletin*, 23(1), pp. 85-101.

Marsh, T. J. (1999). Long River Flow Records. In Herschy, R.W. (ed.). *Hydrometry: Principles and Practice*. John Wiley & Sons Ltd., pp. 161-180.

Martinec, J. and Rango, A. (1989). Merits of statistical criteria for the performance of hydrological models. *Water Resources Bulletin*, 25(2), pp. 421-432.

Masri, S. F., Bekey, G. A. and Safford, F. B. (1980). A global optimization algorithm using adaptive random search. *Applied Mathematics and Computation*, 7, pp. 353-375.

Mathworks (1996). *Matlab – Reference guide*. The Mathworks Inc., Natick, USA.

Mathworks (1993). *Statistics toolbox – User's guide version 2.1.* The Mathworks Inc., Natick, USA.

McCuen, R. H. and Snyder, W. M. (1986). Hydrologic modelling - Statistical methods and applications. Prentice-Hall, USA

McDonnell, J. J., Rowe, L. and Stewart, L. (1999). A combined tracer-hydrometric approach to assess the effect of catchment scale on water flow path, source and age. In Leibundgut, C., McDonnell, J.J. and Schutlz, G. (eds.) Integrated methods in catchment hydrology – Tracer, remote sensing and new hydrometric techniques. IAHS Publication No. 258, pp. 265-273.

McIntyre, N., Lees, M. J. and Wheater, H. S. (2002). Estimation and propagation of parametric uncertainty in environmental models. *J. of Hydroinformatics*, in Review.

McKay, M. D., Beckman, R. J. and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from computer code. *Technometrics*, 21(2), pp. 239-245.

Mehlhorn, J. and Leibundgut, C. (1999). The use of tracer hydrological time parameters to calibrate baseflow in rainfall-runoff modelling. In Leibundgut, C., McDonnell, J.J. and Schultz, G. (eds.) *Integrated methods in catchment hydrology – Tracer, remote sensing and new hydrometric techniques.* IAHS Publication No. 258, pp. 119-125.

Mein, R. G. and Brown, B. M. (1978). Sensitivity of optimized parameters in watershed models. *Water Resour. Res.*, 14(2), pp. 299-303.

Mein, R. G., Laurenson, E. M. and McMahon, T. A. (1974). Simple Nonlinear Model for Flood Estimation. *J. of the Hydraulic Division*, Am. Soc. Civ. Eng., 100(HY11), pp. 1507-1518.

Melching, C. S. (1995). Reliability estimation. In Singh, V.P. (ed.) *Computer models of watershed hydrology.* Water Resources Publishers, USA, pp. 69-118.

Melching, C. S., Yen, B. C. and Wenzel, H. G., Jr. (1990). A reliability estimation in modeling watershed runoff with uncertainties. *Water Resour. Res.*, 26(10), pp. 2275-2286.

Melching, C. S., Yen, B.C. and Wenzel, H. G., Jr., (1991). Output reliability as guide for selection of rainfall-runoff models. Journal of Water Resources Planning and Management, 117(3), pp. 383-398.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines. Journal of Chemistry and Physics, 21, pp. 1087-1091.

Michaud, J. and Sorooshian, S. (1994). Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. *Water Resour. Res.*, 30(3), 593-605.

Mimikou, M. and Gordis, J. (1989). Predicting the mean annual flood and flood quantiles for ungauged catchments in Greece. *Hydrological Sciences J.*, 34(2), pp. 169-184.

Minns, A. W. and Hall, M. J. (1996). Artificial neural networks as rainfall-runoff models. *Hydrological Science J.*, 41, pp. 399-417.

Mohanty, B. P. (1999). Scaling hydraulic properties of a macroporous soil. *Water Resour. Res.*, 35, pp. 1927-1931.

Montesinos, P., Beven, K. J. (1999). Application of a genetic algorithm as sampler selector within the GLUE approach. Poster Presentation – European Geophysical Society Assembly, The Hague, NL.

Moore, R. J. (1985). The probability-distributed principle and runoff prediction at point and basin scales. *Hydrological Sciences Bulletin*, 30(2), pp. 273-297.

Moore, R. J. (1992). Hydrological models - Nonlinear storage models. International Course for Hydrologists, International Institute for Hydraulic and Environmental Engineering, Institute of Hydrology, Wallingford, UK

Moore, R. J. (1999). Real-time flood forecasting systems: Perspectives and prospects. In Casale, R. and Margottini, C. (eds.) *Floods and landslides: Integrated risk assessment*. Springer-Verlag, Berlin, pp. 147-189.

Moore, R. J. and Clarke, R. T. (1981). A distribution approach to rainfall runoff modeling. *Water Resour. Res.*, 17(5), pp. 1367-1382.

Mosley, M. P. (1981). Delimitation of New Zealand hydrologic regions. *J. Hydrol.*, 49, pp. 173-192.

Mous, S. L. J. (1993). Identification of the movement of water in unsaturated soils: the problem of identifiability of the model. *J. Hydrol.*, 143, pp. 153-167.

Mroczkowski, M., Raper, G. P. and Kuczera, G. (1997). The quest for more powerful validation of conceptual catchment models. *Water Resour. Res.*, 33(10), pp. 2325-2335.

Naef, F. (1977). Ein Vergleich von mathematischen Niederschlag-Abfluss-Modellen. In German, Mitteilungen Nr. 26 der Versuchsanstalt fuer Wasserbau, Hydrologie und Glaziologie, Zürich, Switzerland.

Naef, F. (1981). Can we model the rainfall-runoff process today? *Hydrological Sciences Bulletin*, 26(3), pp. 281-289.

Nash, J. E. (1960). A unit hydrograph study with particular reference to British catchments. *Proc. Inst. Civil Eng.*, 17, pp. 249-282.

Nash, J. E. and Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I – A discussion of principles. *J. Hydrol.*, 10, pp. 282-290.

Nathan, R. J. and McMahon, T.A. (1990). Identification of homogeneous regions for the purpose of regionalisation. *J. Hydrol.*, 121, pp. 217-238.

Nathan, R. J. and McMahon, T.A. (1992). Estimating low flow characteristics in ungauged catchments. *Water Resources Management*, 6, pp. 85-100.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimisation. *Computer Journal*, 7, pp. 308-313.

NERC, (1975). *Flood studies report*. 5 Volumes. Natural Environment Research Council, UK.

Niadas, I. A. (1999). Rainfall-runoff model regionalization: An application to Midwest U.S.A. Unpublished M.Sc. Thesis, Department of Civil and Environmental Engineering, Imperial College of Science, Technology and Medicine, London, UK.

Nielsen, S. A. and Hansen, E. (1973). Numerical simulation of the rainfall-runoff process on a daily basis. *Nordic Hydrology*, 4, pp. 171-190.

NWS, (2001). *Calibration of the Sacramento model structure.* http://hsp.nws.noaa.gov/oh/hrl/calb/workshop/parameter.htm, Accessed 15th February (2001).

NWS, (2000). Distributed model intercomparison project. Unpublished Report, National Weather Service Hydrology Laboratory, Silver Spring, MD, USA.

O'Connell, P. E. (1991). A historical perspective. In Bowles, D.S. and O'Connell, P.E. (eds.) *Recent advances in the modelling of hydrologic systems.* Kluwer, Dordrecht, pp. 3-30.

Oreskes, N., Shrader-Frechette, K. and Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263, pp. 641-646.

Osidele, O. O. and Beck, M. B. (2001). Identification of model structure for aquatic ecosystems using regionalized sensitivity analysis. *Water Science and Technology*, 43(7), pp. 271-278.

Ostrowski, M. W. (1992). Ein universeller Baustein zur Simulation hydrologischer Prozesse. *Wasser + Boden* (in German), 11, pp. 755-760.

Øverland, H., and Kleeberg, H.-B. (1993). Moeglichkeiten der Abflussmodellierung unter Nutzung von Geoinformationssystemen. *Mitteilungen des Instituts fuer Wasserwesen*, 45, Neubiberg, (in German).

Özdamar, L. and Demirhan, M. (2000). Experiments with new stochastic global optimization search techniques. *Computers and Operations Research*, 27, pp. 841-865.

Özelkan, E. C. and Duckstein, L. (2000). Multi-objective fuzzy regression: a general framework. *Computers & Operations Research*, 27, pp. 635-652.

Parkin, G., O'Donnell, D., Ewen, J., Bathurst, J. C., O'Connell, P. E. and Lavabre, J. (1996). Validation of catchment models for predicting land-use and climate change impacts. 2. Case study for a Mediterranean catchment. *J. Hydrol.*, 175, pp. 595-613.

Penman, H. L. (1949). The dependence of transpiration on weather and soil conditions. *J. of Soil Science*, 1, pp. 74-89.

Perrin, C. (2000). Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative. Unpublished Ph.D. Thesis (in French), Institut National Polytechnique de Grenoble, Grenoble, France.

Perrin, C., Michel, C. and Andréassian, V. (2000). Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *J. Hydrol.*, 242, pp. 275-301.

Pickup, G. (1977). Testing the efficiency of algorithms and strategies for automatic calibration of rainfall-runoff models. *Hydrological Sciences Bulletin*, 12(2), 257-274.

Pilgrim, D.H. (1983). Some problems in transferring hydrological relationships between small and large drainage basins and between regions. *J. Hydrol.*, 65, pp. 49-72.

Popper, K. (2000). *The logic of scientific discovery.* First published 1959 by Hutchinson education, Routledge, London, UK.

Porter, J. W. and McMahon, T. A. (1971). A model for the simulation of streamflow data from climatic records. *J. Hydrol.,* 13, pp. 297-324.

Post, D. A. and Jakeman, A. (1996). Relationships between physical attributes and hydrologic response characteristics in small Australian mountain ash catchments. *Hydrological Processes*, 10, pp. 877-892.

Post, D. A., Jones, J. A. and Grant, G. E. (1998). An improved methodology for predicting the daily hydrologic response of ungauged catchments. *Environmental Modelling & Software*, 13, pp. 395-403.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flamery, B. P. (1992). *Numerical Recipes in C: The art of scientific computing.* 2nd Edition, Cambridge University Press, UK. (see also http://www.nr.com)

Price, W.L. (1987). Global optimisation algorithms for a CAD workstation. *J. of Optimization Theory and Applications*, 55(1), pp. 133-146.

Pronzato, L., Walter, E., Venot, A. and Lebruchec, J.F. (1984). A general purpose global optimizer: implementation and applications. *Mathematics and Computers in Simulation*, 26, pp. 412-422.

Reed, D. (1999). *Flood estimation handbook – 1: Overview.* Institute of Hydrology, Wallingford, UK.

Refsgaard, J. C. (1997). Parameterisation, calibration and validation of distributed hydrological models. *J. Hydrol.,* 198, pp. 69-97.

Refsgaard, J. C. and Abbott, M.B. (1996). The role of distributed hydrological modelling in water resource management. In Abbott, M.B. and Refsgaard, J.C. (eds.) *Distributed hydrological modelling.* Kluwer Academic Publishers, NL, pp. 1-16.

Refsgaard, J. C. and Knudsen, J. (1996). Operational validation and intercomparison of different types of hydrological models. *Water Resour. Res.,* 32(7), pp. 2189-2202.

Reimer, W. (1990). Estimating hydrological parameters from basin characteristics for large semiarid catchments. In Beran, M.A., Brilly, M., Becker, A. and Bonacci, O. (eds.) *Regionalization in Hydrology.* IAHS Publication No. 191, Wallingford, UK, pp. 187-194.

Rientjes, T. H. M. and Boekelman, R. H. (1998). *Hydrological models.* Unpublished Lecture Notes Cthe443, Civil Engineering Department, Delft University of Technology, Delft, NL.

Rientjes, T. H. M. and Hassanizadeh, M. (2000). Inverse modelling and physically-based rainfall-runoff models. Proceedings (CD) 4th International Conference on Hydroinformatics, Cedar Rapids, IA, USA.

Riggs, H. C. (1990).. Estimating flow characteristics at ungauged sites. In *Regionalisation in Hydrology.* Proceedings of the Ljubljana Symposium, IAHS Publ. No. 191, pp. 159-169.

Rodda, J. C., Downing, R. A. and Law, F. M. (1976). *Systematic Hydrology.* Newness-Butterworths, London, UK.

Robson, A. and Reed, D. (1999). Flood estimation handbook – 3: Statistical procedures for flood frequency estimation. Institute of Hydrology, Wallingford, UK.

Rogers, C. C. M., Beven, K. J., Morris, E. M. and Anderson, M. G. (1985). Sensitivity analysis, calibration and predictive uncertainty of the Institute of Hydrology distributed model. *J. Hydrol.*, 81, pp. 179-191.

Rosenblueth, E. (1981). Two-point estimates in probabilities. *Applied Mathematical Modelling*, 5, pp. 329-335.

Rosenbrock, H. H. (1960). An automatic method for finding the greatest or least value of a function. *Computer Journal*, 3, pp. 175-184.

Rutledge, A. T. (1993). Computer programs for describing the recession of ground-water discharge and for estimating mean ground-water recharge and discharge from streamflow records. Unpublished Report, U.S. Geological Survey, Water Resources Investigations Report 93-4121, Richmond, Virginia.

Savic, D. A. and Pedrycz, W. (1991). Evaluation of fuzzy linear regression models. *Fuzzy Sets and Systems*, 39, pp. 51-63.

Scales, L. E. (1985). *Introduction to non-linear optimization*. Macmillan Publishers Ltd., London

Schakee, J. C. (1999). Model parameter estimation experiment (MOPEX) - Phase II. Unpublished Report.

Scholz, K. (1995). Stochastic simulation of urban hydrological processes. (Stochastische Simulation urbanhydrologischer Prozesse). Unpublished Ph.D. Thesis, Fachbereich Bauingenieur- und Vermessungswesen, Universität Hannover, (in German), Germany.

Schultz, G. A. (1999). A call for hydrological models based on remote sensing, tracers and other modern hydrometric techniques. In Leibundgut, C., McDonnell, J.J. and Schultz, G. (eds.) *Integrated methods in catchment hydrology – Tracer, remote sensing and new hydrometric techniques*. IAHS Publication No. 258, pp. 223-230.

Schütze, M. R., Butler, D. and Beck, M.B. (2002). *Modelling, simulation and control of urban wastewater systems*. Springer-Verlag, London, UK.

Sefe, F. T. and Boughton, W. C. (1982). Variation of model parameter values and sensitivity with type of objective function. *J. of Hydrology (NZ)*, 21(2), pp. 117-132.

Sefton, C. E. M. and Boorman, D. B. (1997). A regional investigation of climate change impacts on UK streamflow. *J. Hydrol.*, 195, pp. 26-44.

Sefton, C. E. M. and Howarth, S. M. (1998). Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. *J. Hydrol.*, 211, pp. 1-16.

Sefton, C. E. M., Whitehead, P. G., Eatherall, A., Littlewood, I. G. and Jakeman, A. J. (1995). Dynamic response characteristics of the Plynlimon catchments and preliminary analysis of relationships to physical catchment descriptors. *Environmetrics*, 6, pp. 465-472.

Seibert, J. (1997). HBV light version 1.2 user's manual. Unpublished Report, Department of Earth Science, University of Uppsala, S.

Seibert, J. (1999a). Conceptual runoff models – fiction or representation of reality? Ph.D. Thesis, University of Uppsala, S.

Seibert, J. (1999b). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology*, 98-99, pp. 279-293.

Seibert, J. (2000). Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences*, 4(2), pp. 215-224.

Seibert, J and McDonnell, J. J. (2002). Multi-criterai calibration of conceptual runoff models – the quest for an improved dialog between modeler and experimentalist. In Duan, Q., Sorooshian, S., Gupta, H.V., Rousseau, A. and Turcotte, R (eds.) *Advances in calibration of watershed models.* AGU Monograph Series, USA, pp. 301-315.

Servat, E. and Dezetter, A. (1993). Rainfall-runoff modelling and water resources assessment in northwestern Ivory Coast. Tentative extension to ungauged catchments. *J. Hydrol.*, 148, pp. 231-248.

Shamseldin, A. Y., O'Connor, K. M. and Liang, G. C. (1997). Methods for combining the outputs of different rainfall-runoff models. *J. Hydrol.*, 197, pp. 203-229.

Shaw, E. M. (1994). *Hydrology in practice.* 3rd Edition, Chapman & Hall, London, UK.

Sherratt, D. J. (1985). Regional soil moisture modelling. Unpublished Ph.D. Thesis, Imperial College of Science, Technology and Medicine, UK

Sherratt, D. J. and Wheater, H. S. (1984). The use of surface resistance - soil moisture relationships in soil water budget models. *Agricultural and Forest Meteorology*, 31, pp. 143-157

Singh, V. P. (ed.) (1995). *Computer Models of Watershed Hydrology.* Water Resources Publishers, USA.

Singh, V. P. and Woolhiser, D. A. (1976). Sensitivity of linear and non-linear surface runoff models to input errors. *J. Hydrol.*, 29, pp. 243-249.

Sorooshian, S. (1981). Parameter estimation of rainfall-runoff models with heteroscedastic streamflow errors – The noninformative data cases. *J. Hydrol.*, 52, pp. 127-138.

Sorooshian, S. and Dracup, J. A. (1980). Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Water Resour. Res.*, 16(2), pp. 430-442.

Sorooshian, S. and Arfi, F. (1982). Response surface parameter sensitivity analysis methods for post-calibration studies. *Water Resour. Res.*, 18(5), pp. 1531-1538.

Sorooshian, S. and Gupta, V. K. (1983). Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness. *Water Resour. Res.*, 19(1), pp. 260-268.

Sorooshian, S. and Gupta, V. K. (1985). The analysis of structural identifiability: Theory and application to conceptual rainfall-runoff models. *Water Resour. Res.*, 21(4), pp. 487-495.

Sorooshian, S. and Gupta, V. K. (1995). Model Calibration. In Singh, V.P. (ed.). *Computer Models of Watershed Hydrology.* Water Resources Publications, USA, pp. 23-68.

Sorooshian, S., Gupta, V. K. and Fulton, J. L. (1983). Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: Influence of calibration data variability and length on mode credibility. *Water Resour. Res.,* 19(1), pp. 251-259.

Sorooshian, S., Duan, Q. and Gupta, V. K. (1993). Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resour. Res.,* 29(4), pp. 1185-1194.

Sorooshian, S., Gupta, H. V., Bastidas, L. (1998). Final report - Calibration of hydrologic models using multi-objectives and visualisation techniques. Unpublished Report, Department of Hydrology and Water Resources, University of Arizona, USA.

Spade, P. V. (2000). Ockham's nominalist metaphysics: Some main themes. In Spade, P.V. (ed.) *The Cambridge companion to Ockham.* Cambridge University Press, Cambridge, UK, pp. 100-117.

Spear, R. C. (1993). Regional sensitivity analysis in environmental systems. In Young, P.C. (ed.) *Concise encyclopaedia of environmental systems.* Pergamon Press, Oxford, UK, pp. 476-479.

Spear, R. C. (1995). Large simulation models: Calibration, uniqueness and goodness of fit. MODSIM'95: International Congress on Modeling and Simulation, Modeling and Simulation Society of Australia, Newcastle.

Spear, R. C. and Hornberger, G.M. (1980). Eutrophication in Peel Inlet – II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research,* 14, pp. 43-49.

Spear, R. C., Grieb, T.M. and Shang, N. (1994). Parameter uncertainty and interaction in complex environmental models. *Water Resour. Res.,* 30(11), pp. 3159-3169.

Srivastava, V. K. and Giles, D.E.A. (1987). *Seemingly unrelated regression equation models: Estimation and inference.* Marcel Dekker Inc., New York, USA.

Steinhorst, R. K. 1979. Parameter identifiability, validation, and sensitivity analysis of large system models. In Innis, G.S. and O'Neill, R.V. (eds.) *Systems analysis of ecosystems.* International Co-operative Publishing House, Fairland, Maryland, USA, pp. 33-58.

Stigter, J. D., Beck, M. B. and Gilbert, R. J. (1997). Identification of model structure for photosynthesis and respiration of algal populations. *Water Science and Technology,* 36(5), pp. 35-42.

Sugawara, M. (1995). Chapter 6: Tank Model. In Singh, V.P. (ed.) *Computer models of watershed hydrology.* Water Resources Publishers, Colorado, USA, pp. 165-214.

Tallaksen, L.M. (1995). A review of baseflow recession analysis. *J. Hydrol.,* 165, pp. 349-370.

Tanakamara, H. (1995). Parameter estimation for the Tank model using global optimization. *Transactions of the JSIDRE,* 178, pp. 103-112.

Tanakamara H. and Burges S. J. (1996). Application of global optimization to parameter estimation of the Tank model. Proceedings of the International Conference on Wa-

ter Resources and Environmental Research: Towards the 21st Century, Kyoto, Japan, 29-31 October 1996, 2, pp. 39-46.

TASC, (1980). Applications of Kalman filtering and maximum likelihood parameter identification to hydrologic forecasting. The Analytic Sciences Corporation, Reading, Massachusetts, USA.

Tasker, G. D. (1982). Comparing methods of hydrologic regionalization. *Water Resources Bulletin*, 18(6), pp. 965-970.

Terano, T., Asai, K. and Sugeno, M. 1992. *Fuzzy systems theory and its applications*. Academic Press Inc., London, UK.

Thiemann, M., Trosset, M. W., Gupta, H. V. and Sorooshian, S. (2001). Recursive Bayesian inference for hydrologic models. *Water Resour. Res.*, 37(10), pp. 2521-2535.

Thornthwaite, C. W. (1948). An approach towards a rational classification of climate. *Geographical Review*, 38, pp. 5-94.

Thornthwaite, C. W. and Mather, J. R. (1955). The water balance. *Publ. Climatol. Lab. Climatol. Drexel. Inst. Technol.*, 8(1), pp. 1-104.

Thyer, M., Kuczera, G. and Bates, B. C. (1999). Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms. *Water Resour. Res.*, 35(3), pp. 767-773.

Todini, E. (1988). Rainfall-runoff modelling – Past, present and future. *J. Hydrol.*, 100, 341-352.

Torrence, C. and Compo, G.P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1), pp. 61-78.

Troendle, C. A. (1985). Variable source area models. In Anderson, M.G. and Burt, T.P. (eds.) *Hydrological forecasting*. John Wiley & Sons Ltd, pp. 347-403.

Troutman, B. M. (1982). An analysis of input errors in precipitation-runoff models using regression with errors in the independent variable. *Water Resour. Res.*, 18(4), pp. 947-964.

Troutman, B. M. (1985). Errors and parameter estimation in precipitation-runoff modelling 1. Theory. *Water Resour. Res.*, 21(8), pp. 1195-1213.

Tung, Y.-K., Yeh, K.-C. and Yang, J.-C. (1997). Regionalization of unit hydrograph parameters: 1. Comparison of regression analysis techniques. *Stochastic Hydrology and Hydraulics*, 11, pp. 145-171.

Uhlenbrook, S. and Leibundgut, C. (1999). Integration of tracer information into the development of a rainfall-runoff model. In Leibundgut, C., McDonnell, J.J. and Schultz, G. (eds.) *Integrated methods in catchment hydrology – Tracer, remote sensing and new hydrometric techniques*. IAHS Publication No. 258, pp. 93-100.

Uhlenbrook, S., Seibert, J., Leibundgut, C. and Rohde, A. (1999). Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structures. *Hydrological Sciences Bulletin*, 44(5), pp. 779-797.

Vandewiele, G. L. and Elias, A. (1995). Monthly water balance of ungauged catchments obtained by geographical regionalization. *J. Hydrol.*, 170, pp. 277-291.

Vandewiele, G. L., Xu, C.-Y. and Win, N.-L. (1991). Regionalization of model types and parameter values of parsimonious monthly rainfall-runoff models. In Tsakiris, G. (ed.) *Advances in Water Resources Technology*. Balkema, Rotterdam, NL, pp. 27-34.

Van Straten, G. and Keesman, K. J. (1991). Uncertainty propagation and speculation in projective forecasts of environmental change: a lake-eutrophication example. *J. of Forecasting*, 10, pp. 163-190.

Vicens, G. J., Rodriquez-Iturbe, I. and Schakee, J. C., Jr. (1975). A Bayesian framework for the use of regional information in hydrology. *Water Resour. Res.*, 11(3), pp. 405-414.

Vitvar, T., Gurtz, J. and Lang, H. (1999). Application of GIS-based distributed hydrological modelling for estimation of water residence times in the small Swiss prealpine catchment Rietholzbach. In Leibundgut, C., McDonnell, J.J. and Schultz, G. (eds.) *Integrated methods in catchment hydrology – Tracer, remote sensing and new hydrometric techniques*. IAHS Publication No. 258, pp. 241-248.

Vogel, R. M., Fernandez, W. and Sankarasubramanian, A. (1999). The calibration and regionalization of a watershed model. Poster Presentation IUGG 1999, Birmingham, UK.

Vrugt, J. A., Gupta, H. V., Bouten, W. and Sorooshian, S. (2003a). A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters. *Water Resour. Res,* in press.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. and Sorooshian, S. (2003b). Effective and efficient algorithm of multi-objective optimization of hydrologic models. *Water Resour. Res,* in press.

Wagener, T. (1998). Developing a prototype knowledge based system to support rainfall-runoff modelling. Unpublished M.Sc. Thesis, Delft University of Technology, Delft, NL.

Wagener, T. and Wheater, H. S. (2002). On the evaluation of conceptual rainfall-runoff models using multiple-objectives and dynamic identifiability analysis. In Littlewood, I. (ed.) *Continuous river flow simulation: Methods, applications and uncertainties*. British Hydrological Society – Occasional Papers, No. 13, Wallingford, UK, pp. 45-51.

Wagener, T., Lees, M. J. and Wheater, H. S. (1999). A generic rainfall-runoff modelling toolbox. *Eos. Trans. AGU*, 80(17), Fall Meeting Supplement, F203.

Wagener, T., Lees, M. J., Wheater, H.S. (2000). Incorporating predictive uncertainty into a rainfall-runoff modelling system. Proceedings of Hydroinformatics 2000 Conference, on CD, Iowa, USA.

Wagener, T., Boyle, D.P., Lees M. J., Wheater, H. S., Gupta, H. V. and Sorooshian, S. (2001). A framework for the development and application of hydrological models. *Hydrology and Earth System Sciences*, 5(1), pp. 13-26.

Wagener, T., Lees, M. J. and Wheater, H. S. (2002a). A toolkit for the development and application of parsimonious hydrological models. In Singh, V.P. and Frevert, D.

(eds.) *Mathematical models of large watershed hydrology – Volume 1.* Water Resources Publishers, USA, pp. 87-136.

Wagener, T., Camacho, L.A. and Wheater, H.S. (2002b). Dynamic parameter identifiability of a solute transport model. *J. Hydroinformatics*, 04.3, pp. 199-211.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S. and Gupta, H. V. (2003a). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes*, 17, pp. 455-476.

Wagener, T., Wheater, H.S. and Gupta, H.V. (2003b). Identification and evaluation of conceptual rainfall-runoff models. In Duan, Q., Sorooshian, S., Gupta, H.V., Rousseau, A. and Turcotte, R. (eds.) *Advances in calibration of watershed models.* AGU Monograph Series, USA, pp. 29-47.

Wagner, B. J. and Harvey, J. W. (1997). Experimental design for estimating parameters of rate-limited mass transfer: Analysis of stream tracer studies. *Water Resour. Res.*, 33(7), pp. 1731-1741.

Wang, Q. J. (1991). The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. *Water Resour. Res.*, 27(9), pp. 2467-2471.

Ward, R. C. and Robinson, M. (2000). *Principles of hydrology.* 4th ed., McGraw-Hill, London, UK.

Wasserman, P. D. (1993). Advances in methods in neural computing. Van Nostrand Reinhold, New York, USA.

Weiler, M., Scherrer, S., Naef, F. and Burlando, P. (1999). Hydrograph separation of runoff components based on measuring hydraulic state variables, tracer experiments, and weighting methods. In Leibundgut, C., McDonnell, J.J. and Schultz, G. (eds.) *Integrated methods in catchment hydrology – Tracer, remote sensing and new hydrometric techniques.* IAHS Publication No. 258, pp. 249-255.

Weisberg, S. (1980). *Applied linear regression.* John Wiley & Sons, Chichester, UK.

Wheater H. S., Bishop K. H. and Beck, M. B. (1986). The identification of conceptual hydrological models for surface water acidification. *Hydrological Processes*, 1, pp. 89-109.

Wheater, H. S., Jakeman, A. J. and Beven, K. J. (1993). Progress and directions in rainfall-runoff modelling. In Jakeman, A.J., Beck, M.B. and McAleer, M.J. (eds.) *Modelling change in environmental systems.* John Wiley & Sons, pp. 101-132.

Whitley, D. (1989). The genitor algorithm and selection pressure: Why rank-based allocation of reproductive trials is best. Paper presented at Third International Conference on Genetic Algorithms.

Wilcox, B. P., Rawls, W. J., Brakensiek, D. L. and Wright, J. R. (1990).. Predicting runoff from rangeland catchments: A comparison of two models. *Water Resour. Res.*, 26(10), pp. 2401-2410.

Wittenberg, H. (1999). Baseflow recession and recharge as nonlinear storage processes. *Hydrological Processes*, 13, pp. 715-726.

Wittenberg, H. and Sivapalan, M. (1999). Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation. *J. Hydrol.*, 102, pp. 29-47.

WMO, (1975). Intercomparison of conceptual models used in operational hydrological forecasting. WMO Operational Hydrology Report No. 7.

Wolock, D. M., Hornberger, G. M. and Musgrove, T. J. (1990).. Topographic effects on flow paths and surface water chemistry of the Llyn Brianne catchments in Wales. *Journal of Hydrology*, 115, pp. 243-259.

Wolpert, D. H. and Macready, W. G. (1995). No free-lunch theorems for search. Santa-Fe Institute working paper 95-02-010.

Wood, E. F. and O'Connell, P. E. (1985). Real-time forecasting. In Anderson, M.G. and Burt, T.P. (eds.) *Hydrological forecasting*. John Wiley & Sons, Chichester, UK, pp. 505-558.

Wood, E .F., Sivapalan, M., Beven, K. and Band, L. (1988). Effects of spatial variability and scale with implications to hydrologic modeling. *J. Hydrol.*, 102, pp. 29-47.

Wood, E. F., Lettenmaier, D. P. and Zatarin, V. G. (1992). A land-surface hydrology parameterisation with sub-grid variability for general circulation models. *J. of Geophysical Research*, 97(D3), pp. 2717-2728.

Wood, W. L. and Calver, A. (1992). Initial conditions for hillslope hydrology modelling. *J. Hydrol.*, 130, pp. 379-397.

Woods, R. A. and Ibbitt, R. P. (1993). Building a rainfall-runoff modeling system: Ideas and practice. In McAleer, M. and Jakeman, A.J. (eds). *International congress on modeling and simulation – Proceedings – Volume I.* Uniprint, University of Western Australia, pp. 25-30.

Woolhiser, D. A. and Brakensiek, D. L. (1982). Hydrologic system synthesis. In Haan, C.T., Johnson, H.P. and Brakensiek, D.L. (eds.) *Hydrologic modelling of small watersheds*. ASAE Monograph, USA, pp. 3-16.

Xevi, E., Christiaens, K., Espino, A., Sewnandan, W., Mallants, D., Sorensen, H. and Feyen, J. (1997). Calibration, validation and sensitivity analysis of the MIKE SHE model using the Neuenkirchen catchment as case study. *Water Resources Management*, 11, pp. 219-242.

Xu, C.-Y. and Singh, V.P. (1998). A review of monthly water balance models for water resource investigations. *Water Resources Management*, 12, pp. 31-50.

Yan, J. and Haan, C. T. (1991). Multiobjective parameter estimation for hydrologic models – Weighting of errors. *Trans. American Society of Agricultural Engineering*, 34(1), pp. 135-141.

Yapo, P. O., Gupta, H. V. and Sorooshian, S. (1996). Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data. *J. Hydrol.*, 181, pp. 23-48.

Yapo, P. O., Gupta, H. V. and Sorooshian, S. (1998). Multi-objective global optimisation for hydrologic models, *J. Hydrol.*, 204, pp. 83-97.

Ye, W., Jakeman, A. J. and Young, P. C. (1998). Identification of improved rainfall-runoff models for an ephemeral low-yielding Australian catchment, *Environmental Modelling & Software*, 13(1), pp. 59-74.

Ye, W., Bates, B. C., Viney, N. R., Sivapalan, M. and Jakeman, A. J. (1997). Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments. *Water Resour. Res.*, 33(1), pp. 153-166.

Yeh, K.-C., Yang, J.-C. and Tung, Y.-K. (1997). Regionalization of unit hydrograph parameters: 2. Uncertainty analysis. *Stochastic Hydrology and Hydraulics*, 11, pp. 173-192.

Young, A. (2002). River flow simulation within ungauged catchments using a daily rainfall-runoff model. In Littlewood, I. (ed.) *Continuous river flow simulation: Methods, applications and uncertainties*. British Hydrological Society – Occasional Papers, No. 13, Wallingford, UK, pp. 31-38.

Young, P. C. (1984). Recursive Estimation and Time-Series Analysis – An Introduction. Springer-Verlag, New York.

Young, P. C. (1985). The instrumental variable method: a practical approach to identification and system parameter estimation. Proceedings of the 7th IFAC/IFORS Symposium on Identification and System Parameter Estimation, York, International Federation of Automatic Control, Vol. 1, pp. 1-15.

Young, P. C. (1992). Parallel processes in hydrology and water quality: A unified time-series approach. *J. of IWEM*, 6, pp. 598-612.

Young, P. C. (2001). Data-based mechanistic modelling and validation of rainfall-flow processes. In Anderson M.G. and Bates P.D. (eds.) *Model validation: Perspectives in hydrological science*. John Wiley & Sons, Chichester, in Press.

Young, P. C. and Beven, K.J. (1994). Data-based mechanistic modelling and the rainfall-flow nonlinearity. *Environmetrics*, 5, pp. 335-363.

Young, P. C. and Beven, K.J. (1991). Comments on computation of the instantaneous unit hydrograph and identifiable component flows to two small upland catchments. *Journal of Hydrology*, 126, pp. 389-396.

Young, P. C. and Beven, K.J. (1990). Comments on Computation of the Instantaneous Unit Hydrograph and Identifiable Component Flows with Application to two Small Upland Catchments by A.J. Jakeman et al. *J. Hydrol.*, 117, pp. 275-300.

Young, P.C. and Jakeman, A.J. (1979a). Refined instrumental variable methods of recursive time-series analysis: Part I. *International J. of Control*, 29, 1-30.

Young, P.C. and Jakeman, A.J. (1979b). Refined instrumental variable methods of recursive time-series analysis: Part II. *International J. of Control*, 29, 621-644.

Young, P.C. and Jakeman, A.J. (1980). Refined instrumental variable methods of recursive time-series analysis: Part III. *International J. of Control*, 30, 741-764.

Young, P. C., Jakeman, A.J. and Post, D.A. (1997). Recent advances in the data-based modelling and analysis of hydrological systems. *Water Science and Technology*, 36(5), pp. 99-116.

Young, P. C., Parkinson, S. and Lees, M.J. (1996). Simplicity out of complexity in environmental modelling: Occam's razor revisited. *J. Applied Statistics*, 23, pp. 165-210.

Yu, P.-S. and Yang, T.-C. (2000). Using synthetic flow duration curves for rainfall-runoff model calibration at ungauged sites. *Hydrological Processes*, 14, pp. 117-133.

Zadeh, L. A. (1963). Optimality and non-scalar-valued performance criteria. *IEEE Transactions on Automatic Control*, 8, pp. 59-60.

Zhao, R. J. and Liu, X. R. (1995). The Xinanjiang model. In Singh, V.P. (ed.) *Computer models of watershed hydrology*. Water Resources Publications, Colorado, USA, pp. 215-232.

Zimmermann, H.-J. (1996). *Fuzzy set theory and its applications*. 3rd Ed., Kluwer Academic Publishers, London, UK.

Index

About the Authors

Thorsten Wagener is a Research Scientist at the National Science Foundation (NSF) Science and Technology Center (STC) for Sustainability of semi-Arid Hydrology and Riparian Areas (SAHRA), located in the Department of Hydrology and Water Resources of the University of Arizona in Tucson, USA. He is currently a German Academic Exchange Service (DAAD) Research Scholar. Dr Wagener received a degree in Civil Engineering from the University of Siegen in Germany and a Masters degree from Delft University of Technology in the Netherlands. Subsequently, he was a Research Assistant at Imperial College London where he received his Ph.D. in 2002. His interests are in particular the building, identification and evaluation of land surface hydrological models, and the development of new modelling and uncertainty analysis procedures. He is also interested in the emerging field of hydroinformatics and in the hydrology of semi-arid areas.

Howard S. Wheater is Professor of Hydrology at Imperial College London and Head of the Environmental and Water Resource Engineering section of the Civil and Environmental Engineering Department. He is a Fellow of the Institution of Civil Engineers (London), Life Member of the International Water Academy (Oslo) and Past-President of the British Hydrological Society. He graduated from the University of Cambridge with a first class degree in Engineering Science and worked in the aeronautical industry before specializing in hydrology and taking his Ph.D. at the University of Bristol in 1977. His research interests include flood hydrology, water resources, water quality and waste management. He also has a particular interest in arid zone hydrology, and has advised overseas government and international agencies on flood and water resource issues throughout the Middle East. In the UK, he is adviser to DEFRA on flood R and D, and chairs the NERC (National Environment Research Council) CEH (Centre for Ecology and Hydrology) review panel on land and water resources. He has recently represented the Government of Hungary at the International Court of Justice concerning the environmental impact of the GNBS Danube dams. He has 30 years' experience of hydrological modelling of surface and subsurface

hydrology and water quality, including in particular issues of uncertainty analysis, model identifiability and, currently, model regionalization.

Hoshin V. Gupta is an Adjunct Professor at the Department of Hydrology and Water Resources of the University of Arizona in Tucson, USA. He is also the Associate Director of SAHRA. He received his Ph.D. in Systems Engineering with specialization in Water Resources Systems from Case Western Reserve University (CWRU) in 1984, and his Masters in Systems Engineering from CWRU in 1982. His Bachelor degree is in Civil Engineering from the Indian Institute of Technology, Bombay, India in 1979. Dr Gupta's primary research interest is in the philosophy, theory and practice of mathematical model building, calibration and application, with particular relevance to land surface hydrologic systems, including rainfall-runoff modelling, flood forecasting, soil-vegetation-atmosphere transfer schemes, hydrochemical modelling and applications of remote sensing in hydrology and climate studies. He is an active researcher in the calibration of complex physical models, and has made significant contributions to the theory and practice of model calibration, global optimisation, and parameter sensitivity analysis. Another interest of his is the theory and application of Artificial Neural Networks.