A TOOLKIT FOR THE DEVELOPMENT AND APPLICATION OF PARSIMONIOUS HYDROLOGICAL MODELS

Thorsten Wagener, Matthew J. Lees, Howard S. Wheater

Department of Civil and Environmental Engineering, Imperial College of Science Technology and Medicine, London, SW19 2BU, UK

ABSTRACT: *A modelling toolkit is described which has been developed to produce parsimonious model structures with a high degree of parameter identifiability. This is necessary if sensible relationships between model parameters and catchment characteristics are to be established, for example for regionalization studies. The toolkit contains two major components. The first is a rainfall-runoff modeling system with a generic architecture of lumped, conceptual or metric-conceptual model elements, which allows alternative model structures to be rapidly constructed and tested. The second component is a Monte-Carlo analysis toolbox combining a number of analysis tools to investigate parameter identifiability, model behaviour and prediction uncertainty. Two example applications are presented. These illustrate the use of multiple objective functions to extract information from a single output time-series for analysis of parameter sensitivity and identifiability, and the trade-off between model complexity and identifiability.*

1. INTRODUCTION

Hydrological models are well-established tools that are widely utilized in engineering practice. The majority of model structures currently used can be classified as conceptual. Adopting the definition given by Wheater et al. (1993), conceptual model structures have two important characteristics: (1) their model structure is specified prior to any modeling being undertaken; and (2) (at least some of) the model parameters do not have a physical meaning, in the sense of being independently measurable, and have to be estimated through calibration against observed data.

Conceptual model structures suffer from a number of problems despite their frequent use and development over a number of decades. One of the major constraints is the lack of parameter identifiability, i.e. different combinations of parameters (e.g. Johnston and Pilgrim, 1976; Beven and Binley, 1992), or even different model structures (e.g. Uhlenbrock et al., 1999) yield similar results in terms of a defined performance measure, or objective function. This results in difficulties in interpreting past behaviour of the catchment system, and hence 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, where no streamflow measurements are available, have to be modelled. One approach to deal with this problem is the regionalization or regional transfer of

parameters of a certain model structure (e.g. Jakeman et al., 1992). A selected model structure is calibrated to a large number of catchments and statistical relationships between the parameters and the characteristics of the catchment such as size, land use or soil types are established. These relationships can then be used to derive parameter values for an ungauged catchment. Uncertainty in the model parameters 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. a high regionalization potential (Lees et al., 1999), is therefore a prerequisite for successful regionalization.

Possible directions of improvement with respect to producing better identified models are: (1) the reduction of model complexity to contain only the components, and therefore parameters, that can be identified from the available data (i.e. parsimonious modeling, e.g. Jakeman and Hornberger, 1993; Young et al., 1996; amongst others), (2) the improved 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., 2000a; amongst others), and (3) the use of additional information, i.e. multi-response data such as water quality data, groundwater levels, or tracer measurements (e.g. de Grosbois et al., 1988; Ambroise et al., 1995; Kuczera and Mroczkowski, 1998; Seibert, 1999; amongst others). It should be noted that the use of additional output variables is unlikely to be particularly useful with respect to regionalization studies, since multi-response data are not commonly available. Therefore this approach is not investigated further here; instead we focus on methods of reducing model complexity and increasing the information that can be retrieved from streamflow measurements.

To this end, a toolkit has been developed to enable the development, analysis and comparison of model structures of different levels of complexity. The aim is to identify the appropriate level of complexity that yields a sufficiently high level of performance, whilst retaining an acceptable level of parameter uncertainty. The toolkit is described, and a limited number of modeling exercises are presented to illustrate its use.

2. PARSIMONIOUS RAINFALL-RUNOFF MODELING

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 above a certain level does not result in significantly improved performance (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 range, lead to acceptable model performance (e.g. Beven and Binley, 1992; Spear, 1995; Kuczera and Mroczkowski, 1998).

Results from previous research in the field of rainfall-runoff modeling suggest that up to five or six parameters can be identified from 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; amongst others). 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 streamflow (e.g. Beven, 1989; Jakeman and Hornberger, 1993; Beck, 1987). These findings have led to the investigation of less complex (parsimonious) model structures that capture the key response modes of the hydrological system (e.g. Hornberger et al., 1985; Jakeman and Hornberger, 1993; Young et al., 1996; amongst others).

The principle of parsimony requires models to have the simplest parameterization that can be used to represent the data (Box and Jenkins, 1976). Parsimonious models have reduced problems of identifiability since only model parameters justified by the evidence, i.e. 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). The principle can be described as a* statement of cautious scientific method *(Spade, 2000). The approach of retaining only necessary components ensures that the model components used are positively affirmed. However, using this approach in the context of rainfall-runoff modeling 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 simple 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 can be observed (Gupta and Sorooshian, 1985; Yapo et al., 1996). Young et al. (1996) describe how the principle of Ockham's razor can be applied to the modeling of environmental systems using a combination of Monte-Carlo techniques, dominant mode analysis, and data-based mechanistic modeling.*

Popper (1983) 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, 1983). In the context of hydrological modeling, it is important to consider that the testability of a model structure will improve in cases where an increasing number of output variables exists that can be compared to measured variables, e.g. predictions of groundwater levels or saturated areas (Nash and Sutcliffe, 1970; Seibert, 1999). Additional information is then available to test potential models, and hence to reject those that show behaviour which does not conform with observations.

The basic assumption underlying this idea is that it is not possible to show that a certain theory (model) is correct, i.e. to validate it, but that it is only possible to say that there is insufficient evidence to refute or falsify the theory. In other words, regardless of how often we see a white swan, we can not conclude that all swans are white, i.e. it is not possible to verify (originating from the Latin word verus *meaning true) the hypothesis that all swans are white. However, the observation of a blue swan would lead to the rejection of this hypothesis, i.e. it can be falsified (Magee, 1973). Different researchers discuss how this concept of falsifiability (Popper, 1983) could be applied in the context of hydrological modeling (e.g. Dooge, 1986; Mroczkowski et al., 1997, Seibert, 1999).*

This evidence suggests that striving for parsimonious model structures is advantageous. It is this aim that has motivated the development of the toolkit described here.

3.   ANALYTICAL MODELING FRAMEWORK

The aim of the analytical modeling framework proposed here is to achieve a balance between the required level of model structural complexity and the complexity that can be supported by the available data (Fig. 1).

A 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 upon the intended modeling 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 modeling purpose.
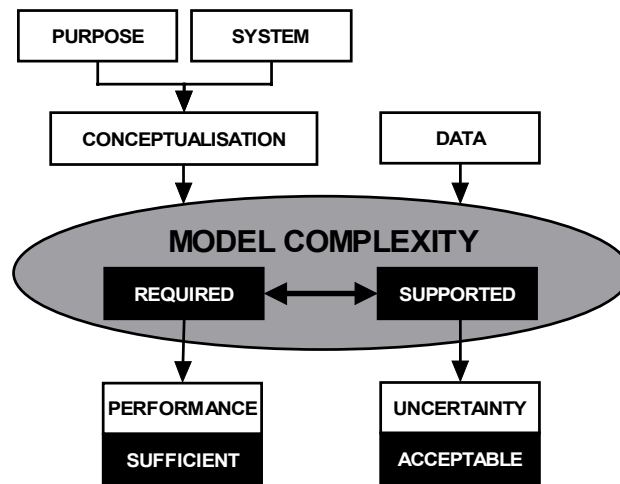

Fig. 1. Proposed analytical framework for model development and application.

The level of structural complexity actually supported by a given data set, is defined here as the number of parameters that can be identified by the information contained within the observed data. Unidentifiable parameters often lead to uncertainty in model predictions, i.e. relatively wide confidence limits, which limit the reliability of predictions and their use in any decision making process. They also limit the suitability of the model structure for purposes such as regionalization as described earlier.

The task therefore is to balance the performance of the model and the identifiability of its parameters such that sufficient prediction accuracy is gained, whilst uncertainty is reduced to an acceptable level.

4. RAINFALL-RUNOFF MODELING COMPONENT

4.1 GENERAL

As noted above, we seek the development of a model structure of appropriate complexity with respect to model performance and associated uncertainty. The philosophy behind this is the recognition that no model structure is suitable for all modeling exercises, but that the appropriate model structure is a function of (Wagener, 1998): (1) the modeling objectives (e.g. required spatial and temporal discretisation, 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 discretization).

An increasing number of modeling shells with different levels of complexity can be found in the literature (for example, Overland and Kleeberg, 1993; Woods and Ibbitt, 1993; Leavesley, 1998; Baalen et al., 1997; amongst others). These systems give their user the possibility to test the suitability of different model components and to combine them in a modular fashion. Components can be modified or added if none of the available components fulfils the problem-specific requirements. The Rainfall-Runoff Modeling Toolbox (RRMT; Wagener et al., 1999) has been developed in particular to produce parsimonious, lumped model structures with a high level of parameter identifiability.

The RRMT is a generic modeling framework or shell that allows its user to implement different model structures in accordance with the framework outlined above. It can be considered therefore to represent a modeling concept, rather than a specific model structure. The RRMT is implemented in the MATLAB (Mathworks, 1996) programming environment.

The model structures that can be implemented are spatially lumped with low or medium levels of complexity (in terms of number of parameters). They can be classified as conceptual or hybrid metric-conceptual in type (Wheater et al., 1993). The latter is related to a systems approach to hydrologic modeling (see examples in Kleissen et al., 1990; Jakeman et al., 1990; Jakeman and Hornberger, 1993). The aim of this approach is to use observations (the metric paradigm) and other prior knowledge to test hypotheses about the structure of component hydrological stores (the conceptual paradigm) at catchment scale (Wheater et al., 1993).

The restriction to a lumped approach could be seen as a limitation since it 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 be valid over a wide range of catchment sizes, and that the aggregated response of a catchment can often be characterized well by such a spatially aggregated model (e.g. Littlewood and Jakeman, 1992; Jakeman and Hornberger, 1993; Jolley, 1995). Blackie and Eeles (1985) point out 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. This stability is required with respect to the spatial distribution of precipitation, vegetation and soil.

4.2 SYSTEM ARCHITECTURE

The RRMT system architecture is based on a modular structure. The modeling component consists of a moisture accounting and a routing module (Fig. 2); 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, broadly representative of the range of current modeling practice, is provided. However, all modules, for example routing or objective function modules, have a specified input and output structure, and can therefore be easily replaced by new or modified modules, so long as they conform to this structure.

The moisture accounting module, which transforms rainfall into effective rainfall, predominantly represents evapotranspiration and associated 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).
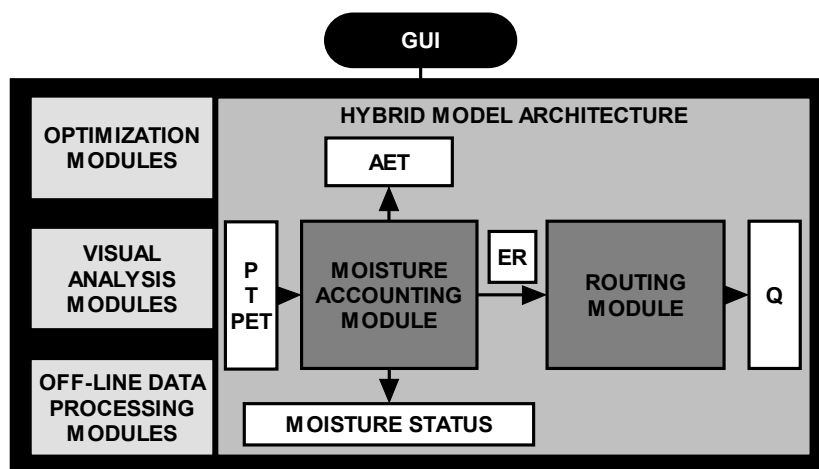

Fig. 2.   System architecture of the Rainfall-Runoff Modeling Toolbox.

A Graphical User Interface (GUI) for the RRMT allows easy access to the toolbox functions. However, since subsequent modeling of a large number of catchments may be required, for example in regionalization studies, it is also possible to run the model from the command line. This allows the user to write a batch file which loads data, changes model settings, runs a calibration, and stores the results for many different cases. Within the batch file, all components of the modeling procedure can be replaced, e.g. the user can try different routing modules or different objective functions for the same data, or perform simulations for different years of data.

4.3 MOISTURE ACCOUNTING MODULES
The moisture accounting module divides the incoming rainfall into losses, through evapotranspiration and associated storage, and water which appears as an output from the catchment system (i.e. effective rainfall). A general water balance equation describing this process can be written in its implicit form as follows (e.g. Hornberger et al., 1998; Blackie and Eeles, 1985)

$$\frac{dV}{dt} = P - Q - AET \qquad (1)$$

where V [L^3] is the volume of water stored in the catchment, P [L^3T^{-1}] is the precipitation rate, Q [L^3T^{-1}] is the rate of surface and subsurface runoff, and AET [L^3T^{-1}] is the rate of actual evapotranspiration.
The moisture accounting module can be represented with varying levels of complexity. Approaches currently implemented in the RRMT (Fig. 3) range from conceptual water balance structures such as those based on Penman's drying curve concept (Penman, 1949; see for example Jolley, 1995) or the Catchment Moisture Deficit (Evans and Jakeman, 1998), to more empirical loss functions related to the Antecedent Precipitation Index (API) such as the Catchment Wetness Index (CWI, Whitehead et al., 1979; Jakeman et al., 1990; Jakeman and Hornberger, 1993). These are typical approaches adopted in parsimonious moisture accounting components, and can be found in many popular model structures.

The CWI is a simple loss function closely related to the well known antecedent precipitation index. The proportion of rainfall r_k, at every time step k, contributing to runoff, i.e. the effective rainfall u_k, is determined by the CWI, in this case based on s_k, which is an indication of the soil moisture state of the catchment, and ranges from zero to one,

$$u_k = \frac{1}{2}(s_k + s_{k-1})r_k \qquad (2)$$

the value for s_k is calculated using the following equation,

$$s_k = cr_k + \left[1 - \frac{1}{\tau(t_k)}\right]s_{k-1}, \qquad (3)$$

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

$$\frac{s_k}{c} = r_k + \frac{s_{k-1}}{c} - \frac{s_{k-1}}{c\tau(t_k)} \qquad (4)$$

This equation 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. 3a.

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

$$\tau(t_k) = \tau_w \cdot \exp[(RT - t_k)f] \qquad\qquad (5)$$

where the reference temperature RT is usually fixed to a nominal value depending on the climate (e.g. RT=10°C for England and Wales (Sefton and Howarth, 1999), RT=20°C for warmer climates (Jakeman et al., 1994)). The two parameters which have to be calibrated are τ_w, the time constant of catchment losses at RT, and the temperature modulation factor f, 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 (Niadas, 1999).

Different modifications of the CWI have been developed to account for specific hydrologic catchment characteristics, for example a power transformation to s_k has been used successfully in catchments with a very flashy response, and the use of a threshold for s_k, below which no rainfall occurs, to enable the modeling of ephemeral rivers (Fig., 3b; Jakeman et al. 1994).

One conceptual water balance component implemented in the RRMT is a version of the Penman model structure used by Jolley (1995) that is based on the Penman drying curve concept (Penman, 1949). The concept assumes that actual evapotranspiration (ae) occurs at the potential (pe) rate while water is available in the root zone or root-reservoir. The root zone is extended by an additional 25mm to allow for capillary rise. The actual rate decreases to a percentage of the potential rate when this soil zone is depleted. The parameterization of this concept 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 water is available. 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, 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% (~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 store (ae_l) is therefore calculated as follows,

$$ae_u = \begin{cases} pe, & D1 < S_{max1} \\ 0, & D1 = S_{max1} \end{cases} \qquad\qquad (6a)$$

$$ae_l = \begin{cases} g(pe - r - S_{max1} + D1), & D1 = S_{max1} \\ 0, & D1 < S_{max1} \end{cases} \qquad\qquad (6b)$$

where pe is the potential evapotranspiration, r is the rainfall input, D1 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 percentage ϕ (usually 15%, Moore, 1999) is bypassed as a direct contribution to rapid groundwater recharge or infiltration excess overland flow,

$$u1 = \begin{cases} \phi(r - pe) & r > pe \\ 0 & r \le pe \end{cases} \qquad (7)$$

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

$$u2 = \max\{(1 - \phi)(r - pe) - D1 - D2; 0\} \qquad (8)$$

where D2 is the soil moisture deficit of the lower store. So that the total effective rainfall u can be calculated as,

$$u = u1 + u2 \qquad (9)$$

The main advantage of the Penman model structure is its efficiency in terms of the number of parameters.

The predominantly vertical processes described by the moisture accounting component usually imply spatial homogeneity of response. This assumption is usually not valid for entire catchments, although good results are obtained in many cases (Blackie and Eeles, 1985; Jakeman and Hornberger, 1993). To address this limitation, Moore and Clarke (1981) introduced a probability distribution approach to storage capacity to account for a heterogeneous catchment response in lumped modeling, e.g. due to the formation of saturated areas in the catchment. A version of this probability distribution loss function is available in the RRMT. Detailed descriptions of this approach are available in Moore and Clarke (1981), or Moore (1999).
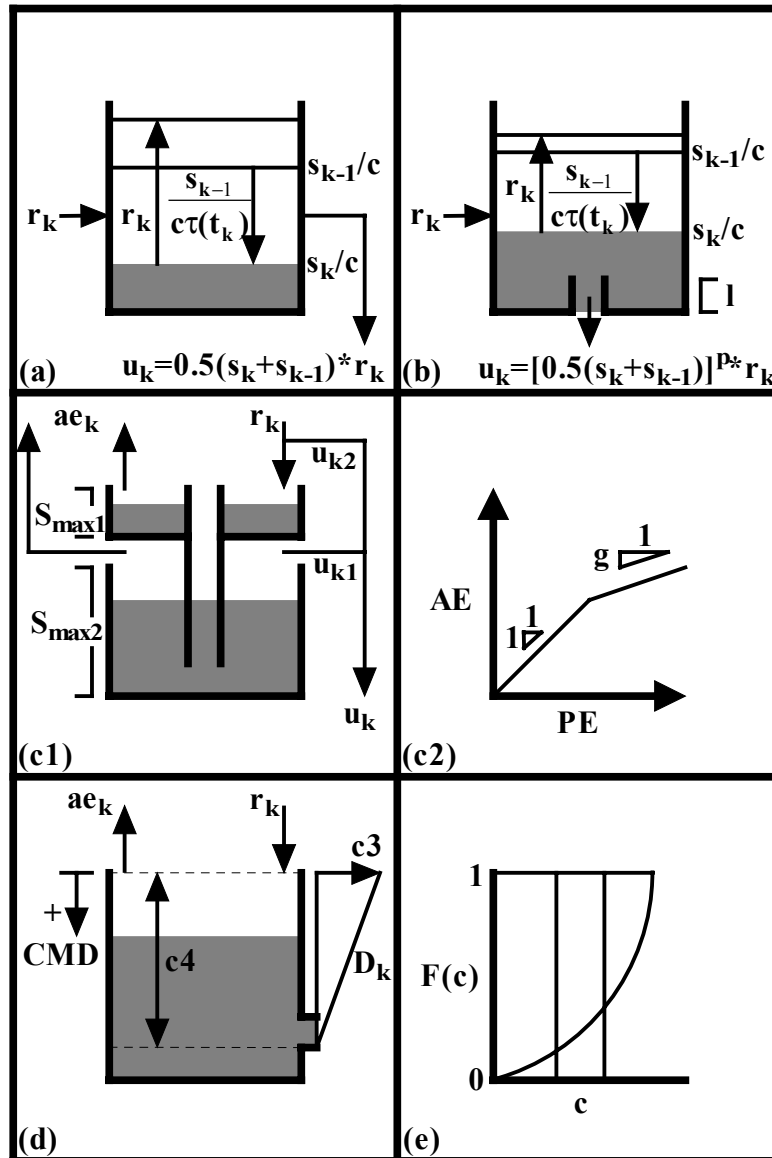
Fig. 3. *Schematic plots of the different moisture accounting modules available: (a) Catchment Wetness Index (CWI), (b) modified CWI, (c1) Conceptual Penman model structure, (c2) Penman drying curve, (d) Catchment Moisture Deficit model structure (CMD), (e) Storage capacity distribution function.*

4.4 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 introduce 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 most commonly used conceptual element 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 two equations.

A storage function can be defined to describe the relationship between outflow of the reservoir and the amount of water stored,

$$S(t) = a \cdot Q^n(t) \qquad (10)$$

where $S(t)$ is the storage [L] at time t, $Q(t)$ is the outflow [L/T] at time t, a is the storage coefficient [$L^{1-n}T^n$], and n is the coefficient of non-linearity [-].
Additionally a mass balance equation describes the change in storage $S(t)/dt$ [L] as the difference between inflow ($u(t)$ [L/T]) and outflow ($Q(t)$ [L/T]) rates,

$$\frac{S(t)}{dt} = u(t) - Q(t) \qquad (11)$$

Jakeman and Hornberger (1993) show that the non-linearity of the relationship between rainfall and runoff can often be accounted for through the consideration of antecedent moisture conditions in a loss function. The consequence is that the remaining transfer from effective rainfall to streamflow can be approximated by a linear relationship.
This leads to the most common form of the conceptual reservoir, the linear reservoir (n = 1). In this case the parameter a becomes the residence time T [T], and the outflow of the reservoir is directly proportional to the storage. The storage function and the mass balance equation can be combined to yield the following model,

$$\frac{dQ}{dt} = \frac{1}{T}[u(t) - Q(t)] \qquad (12)$$

The advantage of the linear reservoir model is computational efficiency since it can be solved analytically.
It can be shown that the linear reservoir is identical to the following first-order discrete-time Transfer Function (TF) (Lees, 2000),

$$Q_t = \frac{b_0}{1 - a_1 z^{-1}} u_t \ ; \ b_0 = \frac{\Delta t}{T} \ ; \ a_1 = 1 - \frac{\Delta t}{T} \qquad (13)$$

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

$$z^{-i} Q_t = Q_{t-i} \qquad (14)$$

Chow et al. (1988) describe 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 nth order single input, single output (SISO) discrete time-system can be written as,

$$Q_t = \frac{B(z^{-1})}{A(z^{-1})} u_{t-\delta} \qquad (15)$$

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

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p} \qquad (16)$$

and

$$B(z^{-1}) = b_0 + b_1 z^{-1} + \cdots + b_m z^{-m} \qquad (17)$$

The TF model structure is therefore described by the triad [p,m+1,δ]. This general model can represent any combination of linear reservoirs connected in parallel and/or series as shown in Fig. 4 (Lees, 2000), 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). A system identification technique that has been successfully used to identify and estimate TF models in the context of rainfall-runoff modeling is the Simple Refined Instrumental Variable technique (SRIV, Young et al., 1996). A TF identified by the SRIV algorithm is one of the routing component options contained in the RRMT.
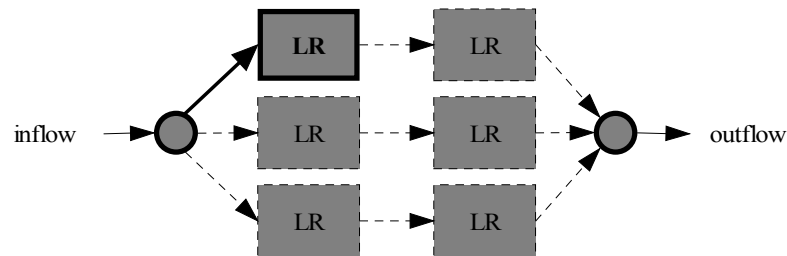

Fig. 4.   *General linear reservoir model.*

The number of reservoir elements required is, amongst other things, dependent upon the modeling 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 modeling (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, 2000b;c). 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 (Hornberger et al., 1999; 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. ephemeral rivers (Jakeman et al., 1994). Physical realism of the selected routing structure is an important criterion if the final aim of the modeling 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 hydraulic head (Hornberger et al., 1998). 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 river bed (Chapman, 1999). However, Wittenberg (1999) found that the behaviour of shallow groundwater reservoirs may be more realistically represented by a non-linear reservoir.

Additional evidence for the need to use non-linear reservoirs in some cases is given by the fact that three parallel linear reservoirs are required to adequately fit the catchment response (Lees and Wagener, 2000b;c). In this case, one store normally represents the quick response, and 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.

The use of a non-linear reservoir is an available option in the RRMT. Linear and non-linear reservoirs can also be combined in parallel or serial structures.
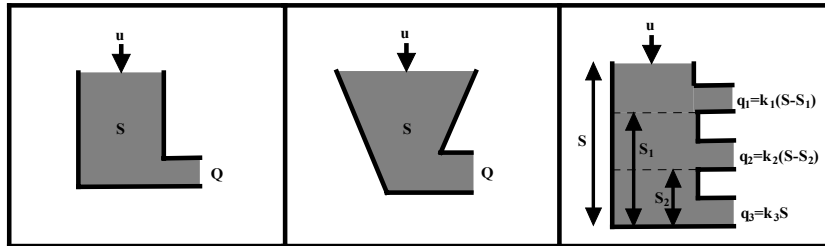

Fig. 5. **Conceptual routing components available in the RRMT: (a) linear; (b) non-linear (example); and (c) leaky catchment structure.**

All 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 leaves some catchments through subsurface pathways. A routing component taking this into account can consist of a linear reservoir with different outlets (Chapman, 1999; see also Sugawara, 1995; and Moore, 1999). The outflow from the bottom outlet is the part of the effective rainfall not contributing to streamflow (Fig. 5).

4.5 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 must be estimated using a calibration procedure whereby the model parameters are adjusted until the system output and model output show an acceptable level of agreement. The agreement is typically measured using an objective function, i.e. some aggregation function of the model residuals, supported by visual inspection of the calculated time series. The model, i.e. a model structure and parameter set combination, that produces the best performance is commonly assumed to be representative of the natural system under investigation.

Most parameters in conceptual rainfall-runoff models define non-linear model equations. The consequence of this is that an iterative search is required to identify the optimum parameter values. This can be done using a manual 'trial-and-error' procedure, by an automatic search algorithm, or by a combination of both approaches (Boyle et al., 2000). Manual calibration is time consuming and difficult in the presence of parameter dependence. However, automatic calibration algorithms can be applied to overcome 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 objective function 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 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. Genetic Algorithms (GA), Wang, 1991; and the Shuffled Complex Evolution algorithm (SCE-UA), Duan et al., 1992), or adaptive random search methods (e.g. Price, 1987). However, various researchers have found that different parameter sets, often widely distributed in the feasible parameter space, lead to similar model performance with respect to a certain objective function (e.g. Johnston and Pilgrim, 1976; Beven and Binley, 1992; etc).

These findings have led some researchers to the conclusion that the idea of 'point identifiability', i.e. a global optimum, is not feasible in the presence of errors in data and model structure, and 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, i.e. parameter sets performing better than a certain threshold. 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 the basis to explore the feasible parameter space in order to identify potential models. Importance sampling based on a uniform prior distribution is most often applied (e.g. Beven and Binley, 1992; Freer et al., 1996), i.e. assuming no prior knowledge of parameter values, other than boundary values. The drawback of this sampling approach is its requirement for a large number of model runs (Spear, 1993).

Gupta et al. (1998) point out that, in the presence of (unavoidable) model structural error, a range of parameter sets is required to adequately simulate all response modes of natural systems. Single parameter sets will favor specific response features, e.g. peak or low flows. This leads to the conclusion that a multi-objective optimization problem exists even for single output model 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, 1985; Jakeman et al., 1990; Young, 1992). 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 the fitting of a large number of different model structures followed by an assessment of the model performance versus parameter identifiability using a 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 modeling (Young, 1992; Young et al., 1996; Lees, 2000). In the rainfall-runoff modeling case this means that a TF structure is only accepted if the TF structure can be interpreted as a combination of linear stores in series and/or parallel.

4.6. OBJECTIVE FUNCTIONS

The performance of a model, i.e. a parameter set and model structure combination, is typically judged using an objective function, usually in combination with visual inspection of the calculated hydrograph. Objective functions aggregate the model residuals, i.e. the part of the observed flow not reproduced by the model, which can be calculated using,

$$\varepsilon(t \mid \theta) = y(t) - \hat{y}(t \mid \theta) \qquad (18)$$

where $\hat{y}(t\,|\,\theta)$ is the calculated flow at time step t using the parameter set θ, $y(t)$ is the observed flow at time step t, and $\varepsilon(t\,|\,\theta)$ is the resulting residual at time step t using parameter set θ.

The task is then to minimize the size of the objective function. A variety of functions are available in the RRMT (Table 1), which can be used to evaluate different aspects of a model's performance.

The most commonly utilized objective functions in hydrological modeling are based on the Simple Least Squares (SLS) function,

$$SLS(\theta) = \sum_{t=1}^{N} \varepsilon(t\,|\,\theta)^2 \qquad (19)$$

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, (3) the residual distribution follows a normal distribution with mean zero.

Table 1.   Objective functions available in the RRMT.

objective function	notation	equation
Nash-Sutcliffe Efficiency[1]	NSE	$NSE(\theta) = 1 - \dfrac{\sum\limits_{t=1}^{N}\left(y(t) - \hat{y}(t\,|\,\theta)\right)^2}{\sum\limits_{t=1}^{N}\left(y(t) - \bar{y}\right)^2}$
Root Mean Square Error	RMSE	$RMSE(\theta) = \sqrt{\dfrac{1}{N}\sum\limits_{t=1}^{N}\left(y(t) - \hat{y}(t\,|\,\theta)\right)^2}$
Heteroscedastic Maximum Likelihood Estimator[2]	HMLE	$\min\limits_{\theta,\lambda} HMLE = \dfrac{\dfrac{1}{N}\sum\limits_{t=1}^{N} w(t)\left(y(t) - \hat{y}(t\,|\,\theta)\right)^2}{\left[\prod\limits_{t=1}^{N} w(t)\right]^{1/N}}$ $w(t) = y(t)^{2(\lambda-1)}$
Bias	BIAS	$BIAS(\theta) = \dfrac{\sum\limits_{t=1}^{N}\left(y(t) - \hat{y}(t\,|\,\theta)\right)}{\sum\limits_{t=1}^{N} y(t)}$
Deviation of Runoff Volumes[3]	DV	$DV(\theta) = \dfrac{\sum\limits_{t=1}^{N} \hat{y}(t\,|\,\theta)}{\sum\limits_{t=1}^{N} y(t)}$
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] American Society of Civil Engineers (1993)
[4] Boyle et al. (2000)

The analysis of the characteristics of the residual distribution is also an important step in evaluating the suitability of a model structure (see for example Yapo et al., 1996; Mroczkowski et al., 1997; or Ljung, 1999 for details). "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., 1999).

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: (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) normal distribution probability scaled plots 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 is described as (Scholz, 1995)

$$r(\tau) = \frac{\sum_{t=1}^{N-\tau} \varepsilon(t \mid \theta) \cdot \varepsilon(t+\tau \mid \theta)}{\sum_{t=1}^{N-\tau} \varepsilon^2(t \mid \theta)} \qquad (20)$$

where τ is a lag ($\tau = 0,1,2,...,N-1$), and $r(\tau)$ ranges from minus one to plus one. The 95% confidence intervals can be calculated as follows,

$$|r(\tau)| < 1.96 \cdot \frac{1}{\sqrt{N}} \qquad (21)$$

where N is the number of data points available. 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 rainfall-runoff modeling applications. Residuals of small temporal discretizations, e.g. daily, are usually related over a number of time steps. Sorooshian and Dracup (1981), and Kuczera (1983) describe consequences and corrective measures if this assumption is violated.

Another problem often encountered in rainfall-runoff modeling 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 objective function (Kottegoda and Rosso, 1997).

A Box-Cox transformation (Kottegoda and Rosso, 1997) which can be written in the following form is a useful transformation in this regard,

$$q(\lambda) = \begin{cases} \dfrac{q^{\lambda} - 1}{\lambda} & for \ \lambda \neq 0 \\ \log(q) & for \ \lambda = 0 \end{cases} \qquad (22)$$

where q is the flow and λ is the transformation parameter.

Sorooshian and Dracup (1981) use this transformation in their Heteroscedastic Maximum Likelihood Estimator (HMLE; Table 1, see also Sorooshian et al., 1993). In this objective function, the parameter λ is estimated simultaneously with the model parameters. 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. Both Box-Cox transformations and the use of the HMLE are available options in RRMT.

Another problem, already briefly indicated, occurs in terms of selection of an appropriate objective function when automatic search algorithms are applied. The drawback of single-criterion algorithms is that the calibration result is fully dependent on one objective function (Gupta et al., 1998; Boyle et al., 2000). This can lead to an unreasonable emphasis on the fitting of a certain aspect of the response, e.g. peak flows, whilst neglecting the model performance with regard to another aspect, e.g. low flows (Fig. 6). 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, therefore constraining the usefulness of automatic calibration (Boyle et al., 2000). A multi-criteria approach is proposed by Gupta et al. (1998) to address this problem. The objective of this approach 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 (Wheater et al., 1986; Wagener et al., 2000a); and (3) to assess the suitability of the model structure to represent the natural system, i.e. identify model structural insufficiencies (Gupta et al., 1998; Boyle et al., 2000). Or as Yan and Haan (1991) state "...the use of multiple objective criteria for parameter estimation permits more of the information contained in a 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".
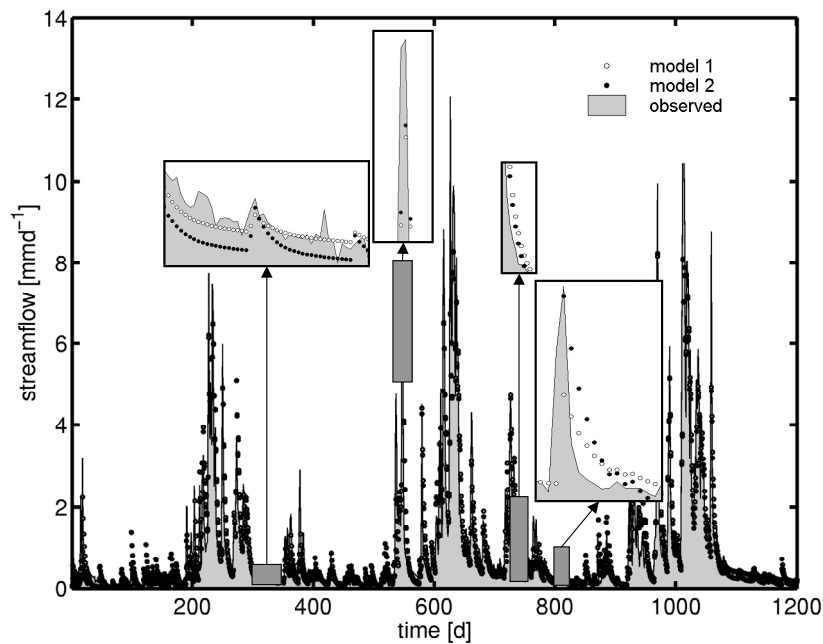
Fig. 6. *Plot showing predictions using two different parameter sets with an identical model structure. Both realizations yield similar values of the Nash-Sutcliffe Efficiency measure (0.82), but show differences in fit when the response is analyzed closely.*

A simple application of multi-objective optimization is the definition of specific objective functions for water resources management. Often, a specific flow range, e.g. between the minimum environmentally acceptable flow and a maximum water supply abstraction rate, is of particular interest. Specifying an objective function 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, 2000b;c). Measures available in the RRMT allow the calculation of the RMSE above (FH) and below (FL) a user specified threshold, and within a certain flow range (FM), see Table 1.

Other multiple objectives that can be selected are the RMSE based on a modified segmentation scheme suggested by Boyle et al. (2000, see application example for details), or the RMSE for a warming up period to estimate initial conditions, FWU. The RRMT allows users to add objective functions in a modular way depending on an application's requirements.

4.7 VISUAL ANALYSIS MODULES
Different plotting options are available in RRMT to analyze 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; (3) flow duration curves and volumetric fit, etc. These plots are helpful tools that enable an assessment of model performance from different perspectives.

4.8 SUMMARY
In summary, the RRMT is a generic modeling shell that allows its user to implement, evaluate and modify lumped, conceptual or hybrid metric-conceptual model structures. A variety of moisture accounting and routing components are provided and the addition of new structures is straightforward. Different data manipulation, optimization, and visualization options are available to calibrate and evaluate the model. The implemented approach is in accordance with the requirements outlined in the analytical framework described earlier in the text.

5. MONTE-CARLO ANALYSIS TOOLBOX

5.1 GENERAL

The detailed investigation of model performance in terms of parameter sensitivity and identifiability, the suitability of a particular model structure, and prediction uncertainty, are increasingly important parts of the modeling task. The understanding of model behaviour and performance gained increases the transparency of the modeling procedure and helps to assess the reliability of modeling results.

The Monte-Carlo Analysis Toolbox (MCAT, Wagener et al., 1999; Lees and Wagener, 2000a) includes a number of analysis methods to evaluate the results of Monte-Carlo parameter sampling experiments or model optimization methods based on population evolution techniques, for example, the Shuffled Complex Evolution algorithm (SCE-UA; Duan et al., 1992). Functions contained in the MCAT include an extension of the Regional Sensitivity Analysis (RSA, Spear and Hornberger, 1980; Hornberger and Spear, 1980) proposed by Freer et al. (1996), various components of the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992; Freer et al., 1996), options for the use of multiple-objectives for model assessment (Gupta et al., 1998; Boyle et al., 2000), response surface plots, and a measure to evaluate parametric identifiability for a selected objective function or in a dynamic fashion.

5.2 SYSTEM ARCHITECTURE

The MCAT is a collection of MATLAB (Mathworks, 1996) analysis and visualization functions integrated through a graphical user interface (Fig. 7). It can also be accessed through an interface from the RRMT. Note that the MCAT is not specifically related to rainfall-runoff modeling and can be used to analyze the results of any dynamic mathematical model.
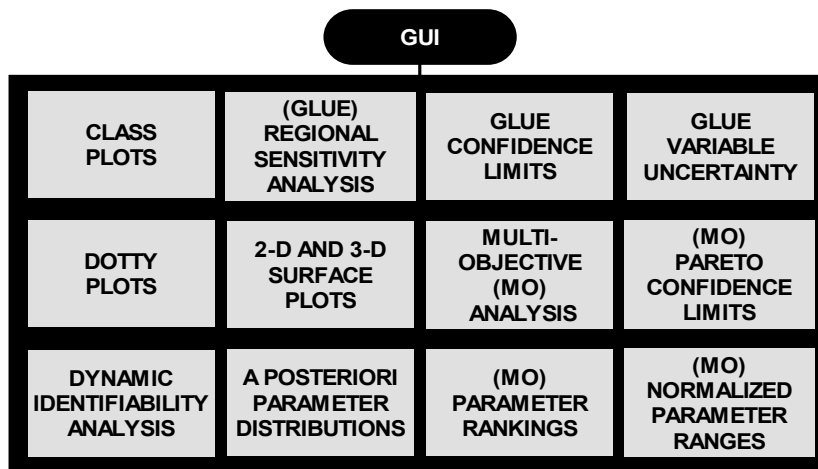

Fig. 7.   System architecture of the Monte-Carlo Analysis Toolbox.

5.3 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 dimension 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 analyzes the sensitivity of different parameters

without referring to a certain point in the parameter space, e.g. the most likely value for a specific parameter (Spear, 1993).
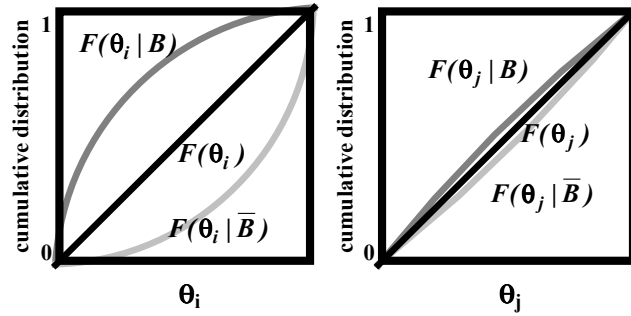

Fig. 8. *Cumulative distributions of initial ($F(\theta_x)$), 'behavioural' ($F(\theta_x | B)$) and 'non-behavioural' ($F(\theta_x | \overline{B})$) populations for a sensitive parameter θ_i, and a (conditionally) insensitive parameter θ_j.*

The RSA method starts with the Monte-Carlo sampling of N points in the feasible parameter space based on a uniform distribution. The resulting parameter population is partitioned into a behavioural (B) and a non-behavioural ($\overline{B}$) group. This division can for example be based on the predicted state of the system (Spear and Hornberger, 1980) or on a measure of performance (Hornberger et al., 1985; Beven and Binley, 1992, amongst others). The cumulative distributions for both groups ($F(\theta_x | B)$ and $F(\theta_x | \overline{B})$) are computed (Fig. 8). A separation between the curves indicates that the initial distribution $F(\theta_x)$ is divided by the classification into behavioural and non-behavioural. This indicates that the parameter θ_x is sensitive, i.e. has an important effect on the model result. The significance of the separation can be estimated using statistical tests such as the Kolmogorov-Smirnov-two-sample test (Kottegoda and Rosso, 1998), and a heuristic ranking scheme can be introduced (Spear and Hornberger, 1980).

However, the lack of separation between the cumulative distributions of $F(\theta_x | B)$ and $F(\theta_x | \overline{B})$ is only a necessary, but not a sufficient condition for insensitivity of θ_x (Spear, 1993). It is possible that it is 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 one or different objective functions.

A modification of the RSA approach proposed by Freer et al. (1996) is implemented in the MCAT to visually inspect the sensitivity of the different parameters with respect to the selected objective function. Freer et al. (1996) split the parameter population, ranked on their objective function 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 indicate parameter sensitivity. Splitting the parameter population into ten groups, instead of just dividing it into behavioural and non-behavioural parameter sets as in the original method, avoids the selection of a threshold value, and increases the information gained by the analysis.

The 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).
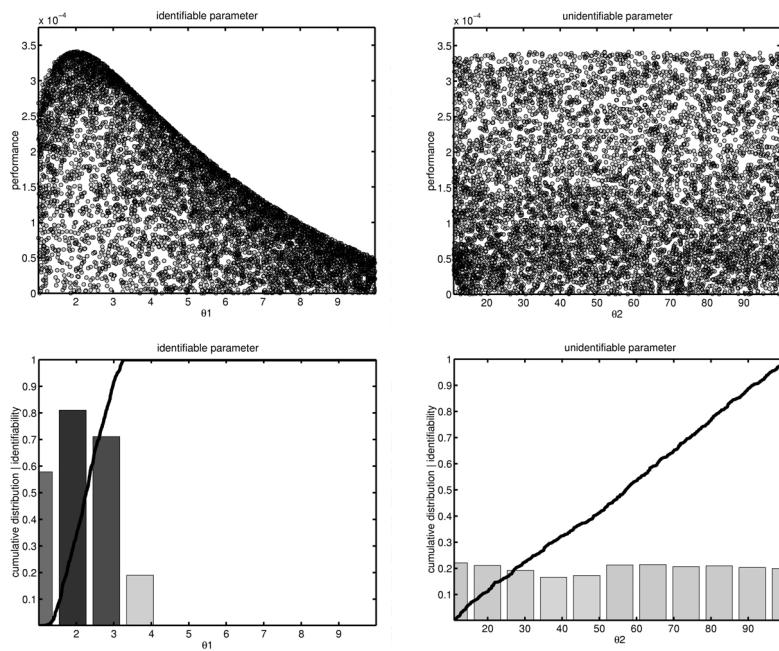
Fig. 9. Example of a well identified and a poorly unidentified parameter. The top row shows scatter plots of parameter versus measure of performance. It has to be considered that these projections into a single parameter dimension can, however, hide some of the structure of the response surface (Beven, 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.

However, parameter sensitivity is only 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 conceptual rainfall-runoff 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 objective functions that aggregate the response vector into a single value, often limit the success of parameter identification as described earlier.

The RSA procedure outlined above has been extended to investigate the identifiability of a parameter. 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 approximate 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. Fig. 9 shows how the gradients for a well identified and a poorly unidentified parameter are distributed. An example of the utility of this measure is included as part of the demonstration application described later. The identifiability measure can also be used in a dynamic way to link parameters, and therefore model components, to system response modes in an objective fashion. The flow-chart in Fig. 10 describes the steps required to perform the dynamic identifiability analysis. In short, the

identifiability measure is calculated as a function of the residuals over a period of X days before and after the time step analyzed. The size of X is dependent upon the number of time steps over which the parameter is influential.
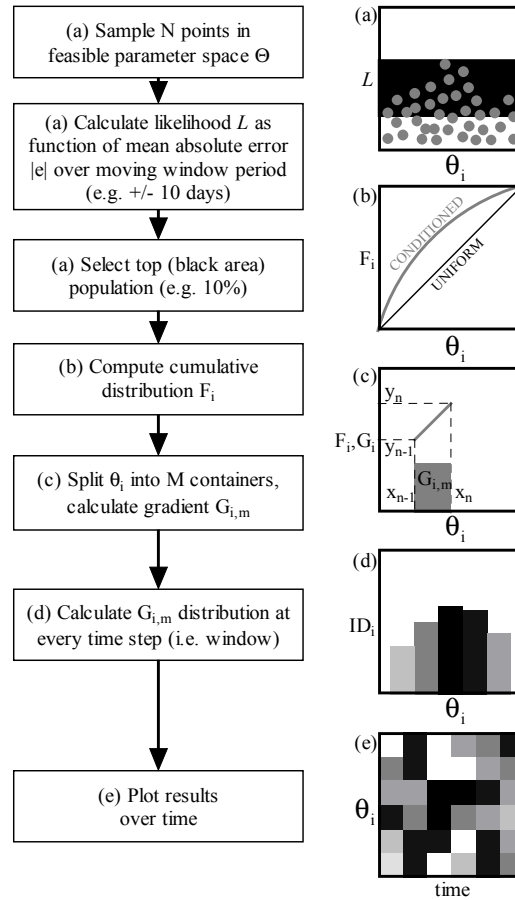

Fig. 10. Dynamic identifiability analysis flowchart.

5.4 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 objective function to measure the model performance with respect to different system responses, and analyzing the variation in the resulting parameter populations can be used to estimate the degree of structural uncertainty present. Multi-objective plots, i.e. scatter plots of different objective functions 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 show a trade-off front, described by the Pareto set (Gupta et al., 1998). A parameter set is termed Pareto optimal if, by changing its parameters, an improvement with respect to one objective function results in the degradation with respect to another objective function (Chankong and Haimes, 1993). 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 best parameter sets for different objective functions on a normalized scale to see the variation

caused by the different measures (for more details of multi-objective optimisation see Sorooshian et al., 2000, 'Emerging paradigms for the calibration of hydrologic models', in Vol.1).

5.5 PREDICTION UNCERTAINTY
Uncertainties in model parameters propagate to prediction uncertainties. 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, i.e. since 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 from a uniform distribution. A simulation is then performed using each parameter set and a 'likelihood' measure to evaluate the performance of the sets 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. The likelihood values of parameter sets below a certain, user selected threshold, are set to zero, since they are considered to be 'non-behavioural'. The likelihood values of the retained models are use to weight the predictions calculated with the corresponding parameter set. The uncertainty in the parameter values, and implicitly the uncertainty in the data, is propagated into the model response in the form of confidence limits of 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.

5.6 SUMMARY
In summary the MCAT aims to provide a tool which addresses the emerging requirement to include detailed investigations of model behaviour, performance, and prediction uncertainties as an integral part of the modeling 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 analyze the role of multiple-objectives (including the option to calculate the Pareto optimal solution and plotting the response range produced by it). It also allows the estimation of dynamic parameter identifiability.

6. APPLICATION EXAMPLES

To illustrate the use of the toolboxes, application examples are presented based on the River Medway catchment (1256.1km^2), which is located in South East England. Almost seven years of daily naturalized flows, precipitation, potential evapotranspiration and temperature data are available for use in the two modeling exercises presented.
The model structures used are based on the Penman two-store model and different combinations of linear and non-linear conceptual reservoirs. This model structure has been found to be appropriate for UK catchments such as the Medway that are 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 (1989-96)).

6.1 USE OF MULTIPLE OBJECTIVE FUNCTIONS
The first example demonstrates how the importance of different components of a simple model structure varies with different response modes of the catchment system. The model structure used is the Penman two-store model, connected to two linear reservoirs in parallel. A uniform random search (10000 samples) is used to investigate the feasible parameter space.

Different response modes are identified using a partitioning scheme suggested by Boyle et al. (2000). This approach is based on the assumption that the behaviour of the catchment is different during periods 'driven' by rainfall and periods without rain, i.e. 'non-driven'. Further, the periods classified as non-driven can be distinguished 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 approach, developed for a semi-arid catchment by Boyle et al. (2000), is slightly modified here to render it suitable for temperate catchments.
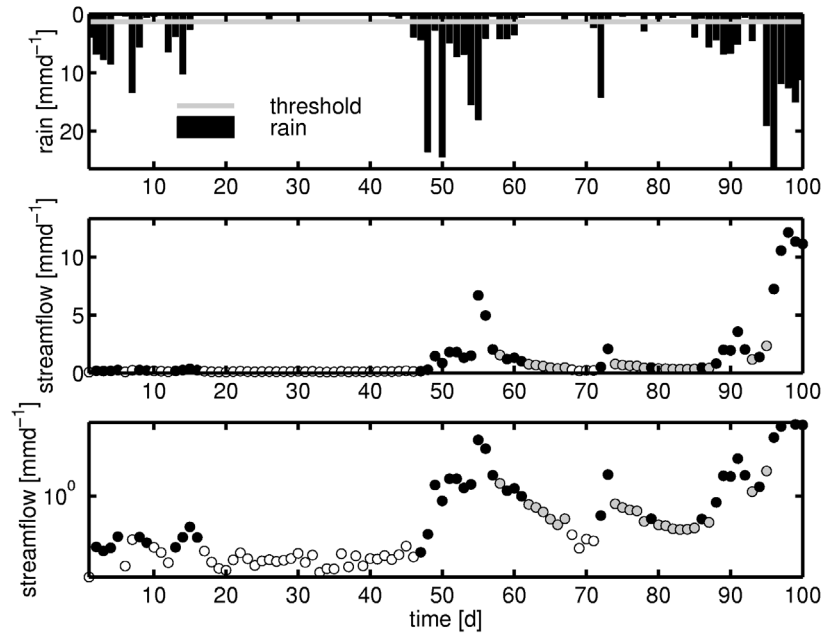

Fig. 11. *A hundred day period of Medway streamflow and rainfall segmented using the scheme modified from Boyle et al. (2000). The dots indicate time steps classified as driven (black), non-driven quick (gray), and non-driven slow (white).*

The time steps corresponding to each period are identified through an analysis of the precipitation data and the time of concentration of the catchment (Fig. 11). The time steps with rainfall larger than a certain threshold (e.g. mean of the square-root of the rainfall), lagged by the time of concentration of the catchment, are classified as driven. 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, and the rest are classified as non-driven-quick. The model performance during these three periods (QD, QQ, and QS) is estimated by calculating the RMSE (FD, FQ, FS) over each period.
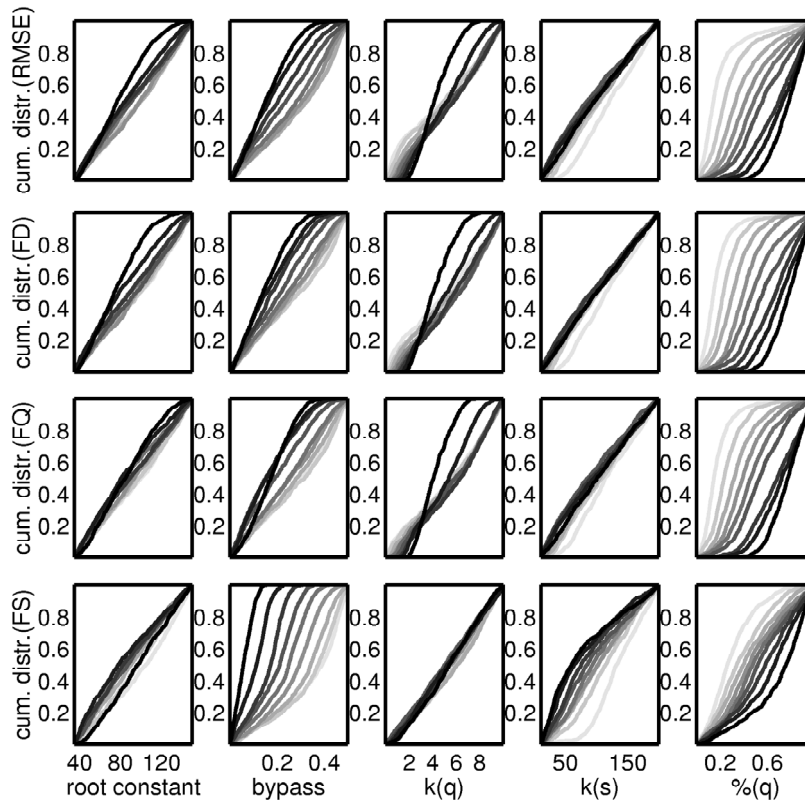
Fig. 12. RSA plot showing that the sensitivity of different parameters varies with the response period of the system using the different objective functions RMSE, FD, FQ, and FS. Darker lines indicate better performing models.

The RSA approach described earlier can be used to investigate the response mode specific character of the parameter sensitivity (Dunn, 1999; Wagener et al., 1999). Fig. 12 shows how the sensitivity of the different parameters varies with the response mode of the system even for this simple model structure. The top row shows the parameter sensitivity with respect to the RMSE over the complete calibration period, while the subsequent rows show the sensitivity during the different response modes as described above. The overall RMSE and FD result in almost identical sensitivities, while FQ and especially FS show different behaviours. The bypass (fraction of rainfall contributing directly to runoff) and the residence time of the linear reservoir responsible for routing the slow response reveal increased sensitivities during the non-driven slow period. The influence of the quick response residence time on the other hand is very low. The percentage effective rainfall routed through the quick response reservoir, %(q), is the only parameter that shows a similar sensitivity during all response modes.

The measure of identifiability introduced earlier in the text, i.e. the (segmented) gradient of the cumulative parameter distribution for the best performing group, can be used here to demonstrate that the change in the identifiability of different parameters is similar to that of their sensitivity (Fig. 13). The results show that the root constant parameter shows similar levels of identifiability for all periods, but the estimated parameter value for the slow response differs considerably from that for the other objective functions. This suggests limitations caused by the simple model structure, compensated by changes in the parameter value to fit different aspects of the system response. The bypass and k(q) show higher levels of identifiability for specific measures indicating that these parameters are identifiable mainly during certain periods. These are periods of slow response for the bypass and the

remaining periods for k(q). More information about the parameters is available during these periods. Parameter %(q) is identifiable during all periods. Additionally all values are in close proximity indicating that this parameter is well identified, and stable during all response modes. Parameter k(s) shows only a slight improvement in identifiability during slow response periods. The identifiability values are relatively low indicating identifiability problems with respect to this parameter.

This perceptual segmentation scheme can be supported by a dynamic identifiability analysis. Fig. 14 shows the variation in identifiability over time for the bypass parameter as an example. A 20 day moving window is used to estimate an identifiability value at every time step. The result shows that this parameter is highly identifiable during summer periods. During these periods a low bypass value is required to avoid erroneous prediction of hydrograph peaks at the end of the summer.
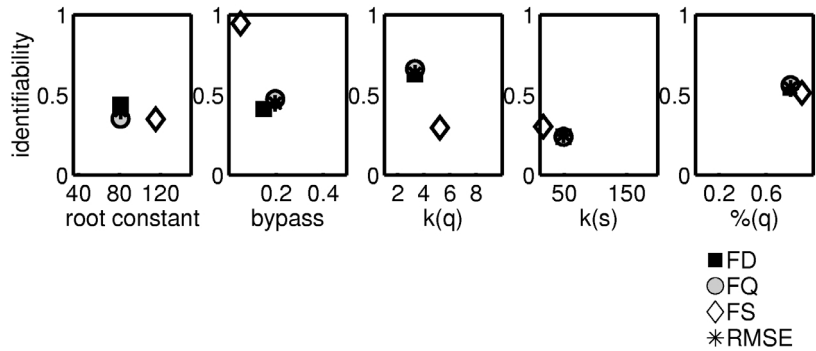

Fig. 13. Plot showing the location of the maximum identifiability value for the different model parameters using the objective functions RMSE, FD, FQ, and FS.
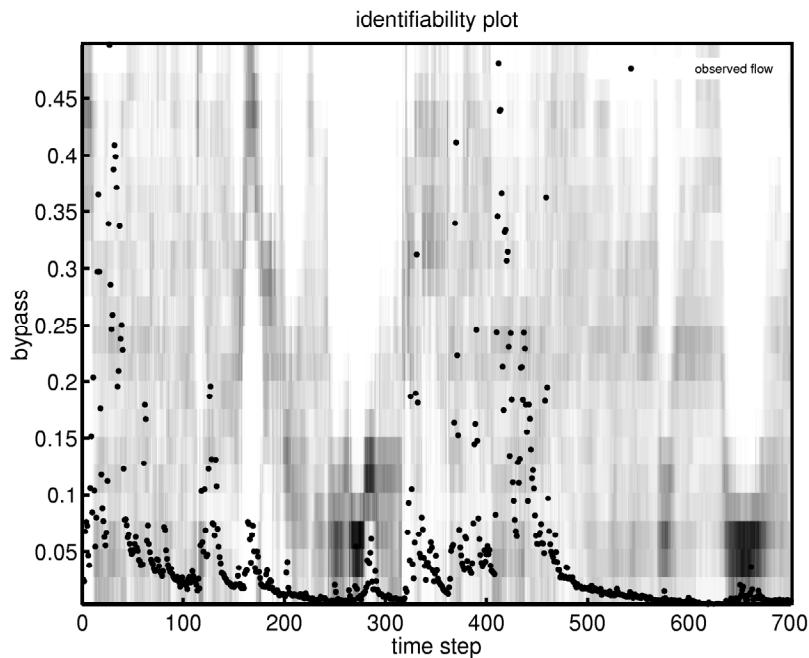

Fig. 14. Dynamic identifiability plot for the bypass parameter.

This example demonstrates that the use of multiple objective functions can lead to an increase of information with respect to parameter sensitivity and identifiability. A multi-objective framework can therefore lead to improved understanding of model behaviour during different response modes, and indicate limitations in the model structure, as evidenced by the varying parameter values observed.

6.2 COMPLEXITY VERSUS IDENTIFIABILTY

The second example compares model structures of different levels of complexity with respect to performance and identifiability. The model structural complexity is increased from a simple bucket, emptied by overflow and/or potential evapotranspiration, to a two-store model based on the Penman drying curve concept described earlier (see Table 2). Additionally the routing component is varied using different combinations of linear and non-linear stores. A model structure is classified as more complex when the number of free parameters is higher or the structural complexity is increased, i.e. a change from bucket to Penman structure. For simplicity, no distinction is made between linear and non-linear approaches.

Table 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 performance of each model structure is evaluated using the Nash-Sutcliffe Efficiency (NSE, Nash and Sutcliffe, 1970, Table 1). However, the NSE criterion favors models that are able to reproduce the catchment response at high flows (Legates and McCabe, 1999). An improvement in performance through increased model complexity, for example during periods of recession, could therefore go unnoticed. This can be avoided by using the partitioning scheme already applied in the first example, and calculating the weighted addition of the NSE for each period. 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,

$$PEF = \frac{1}{3} NSE_D + \frac{1}{3} NSE_Q + \frac{1}{3} NSE_S \qquad (23)$$

where NSE_x is the performance measure during driven (QD), non-driven quick (QQ), and non-driven slow (QS) periods. A MATLAB implementation of the SCE-

UA approach (Duan et al., 1992) is used to optimize the model performance with respect to the defined objective function.

The identifiability of each model structure is investigated using the measure described earlier in the text. A Monte Carlo sampling procedure (10000 samples) based on a uniform distribution 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, is calculated. The mean of all parameter 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 analyzing the stability of the identifiability measure while varying the size of the parameter population considered to calculate it.

Plotting the measures of performance and identifiability against the model structures leads to the graph shown in Fig. 15. The result suggests that increased model complexity leads to increased performance while the identifiability of the model parameters decreases. A trade-off between these two features is required if, for example, it is intended that the model structure is to be use within a regionalization project. The decline in performance with regard to model M-7 is caused by the more suitable slow flow component of model M-6 (a non-linear store).
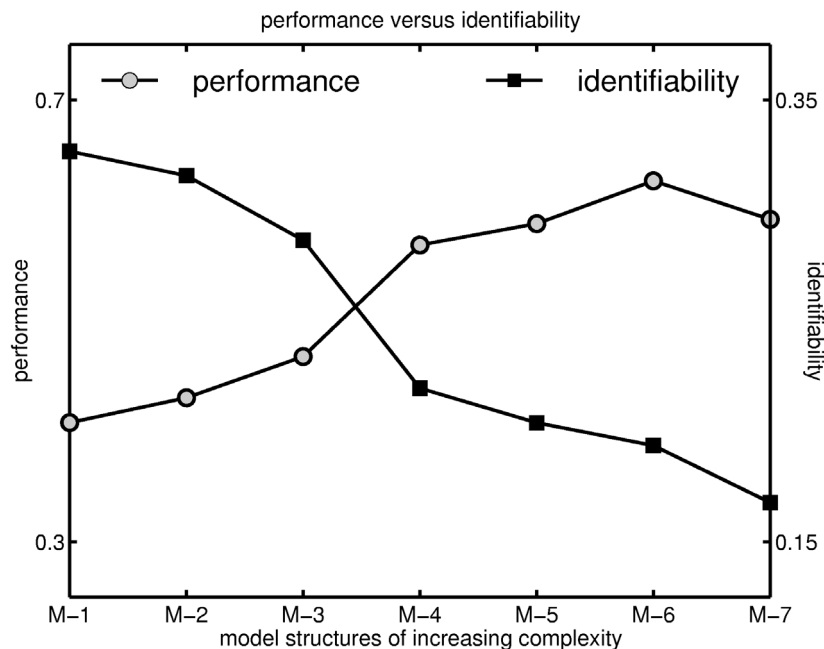

Fig. 15. Performance versus identifiability plot for the different model structures investigated.

7. CONCLUSIONS

The toolkit presented here facilitates the development and analysis of lumped, parsimonious model structures using state-of-the-art modelling techniques. The RRMT allows the implementation of conceptual, or hybrid metric-conceptual model structures. Its major advantage is a high degree of structural flexibility which allows the quick implementation and evaluation of different model structures to identify the most suitable one for the circumstances at hand. The MCAT enables the detailed investigation of model performance, parameter sensitivity and identifiability, model structure suitability, and prediction uncertainty. The toolkit has been developed in accordance with the analytical framework suggested in chapter 3 (Fig. 1). A

tradeoff between the required levels of model complexity in terms of model performance, and the supported levels of model complexity, in terms of identifiability of the model's parameters, can be made to identify the most suitable model structure for a certain modelling application. For instance, if regionalization is required, models with highly identifiable parameters should be selected at the expense of some loss of performance.

Both Matlab toolboxes described here are freely available for non-commercial use from our web-site, see http://ewre.cv.ic.ac.uk/.

REFERENCES
Akaike, H. 1974. A new look at statistical model identification. IEEE Trans. Auto. Control, ***AC-19****, 716-723.*

Ambroise, B., Perrin, J.L. and Reutenauer, D. 1995. Multicriterion validation of a semidistributed conceptual model of the water cycle in Fecht catchment (Vosges massif, France). Wat. Resour. Res.*, **31(6)***, 1467-1481.*

ASCE 1993. Criteria for evaluation of watershed models. J. Irrigation Drainage Eng.*, **119(3)***, 429-442.*

Baalen, S.J.A. van, Stoppelenburg, F.J. and Garritsen, A.C. 1997. Kennisoversicht instrumentarium verdrogingsbestrijding, NOV-rapport 13.1, RIZA-rapport 97.019, (in Dutch), NL.

Bastidas, L.A. 1998. Parameter estimation for hydrometeorological models using multi- criteria methods. Ph.D. Diss., Dep. Hydrol. Wat. Resour., The University of Arizona, USA.

Beck, M.B. 1987. Water quality modeling: A review of the analysis of uncertainty. Wat. Resour. Res.*, **23(8)***, 1393-1442.*

Beven, K.J. 1989. Changing ideas in hydrology – The case of physically-based models. J. Hydrol.*, **105***, 157-172.*

Beven , K.J. 1998. Generalized Likelihood Uncertainty Estimation (GLUE) - User manual. University of Lancaster, UK.

Beven, K.J. and Binley, A.M. 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrol. Proc.*, **6***, 279-298.*

Blackie, J.R. and Eeles, C.W.O. 1985. Chapter 11: Lumped catchment models. In Anderson, M.G., and Burt, T.P. (eds.) Hydrological forecasting. John Wiley & Sons Ltd., UK, 311-345.

Box, G.P. and Jenkins, G.M. 1976. Time series analysis: Forecasting and control. Holden-Day, San Francisco, USA.

Boyle, D.P., Gupta, H.V. and Sorooshian, S. 2000. Towards improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Submitted to Wat. Resour. Res.

Brazil, L.E. and Hudlow, M.D. 1981. Calibration procedures used with the National Weather Service Forecast System. In: Haimes, Y.Y. and Kindler, J. (eds.) Water and related land resources systems. Pergamon, New York, USA, 457-466.

Calder, I.R., Harding, R.J., Rosier, P.T.W. 1983. *An objective assessment of soil-moisture deficit models*. J. Hydrol., **60**, 329-355.

Chankong, V. and Haimes, Y.Y. 1993. *Multiobjective optimization: Pareto optimality. In Young, P.C. (ed.) Concise encyclopedia of environmental systems. Pergamon Press, Oxford, UK*, 387-396.

Chapman, T. 1999. *Comparison of algorithms for stream flow recession and baseflow separation*. Hydrol. Proc., **13**, 701-714.

Chow, V.T. and Kulandaiswamy, V.C. 1971. *General hydrologic system model*. J. Hyd. Div. - Am. Soc. Civ. Eng., **97(HY6)**, 791-804.

Chow, V.T., Maidment, L.W. and Mays, L.W. 1988. Applied hydrology. *McGraw-Hill, USA*.

de Grosbois, E., Hooper, R.P. and Christophersen, N. 1988. *A multisignal automatic calibration methodology for hydrochemical models: A case study of the Birkenes model*. Wat. Resour. Res., **26**, 1299-1307.

Dooge, J.C.I. 1986. *Looking for hydrologic laws*. Wat. Resour. Res., **22(9)**, 46S-58S.

Duan, Q., Gupta, V.K. and Sorooshian, S. 1992. *Effective and efficient global optimization for conceptual rainfall-runoff models*. Wat. Resour. Res., **28(4)**, 1015-1031.

Duan, Q., Gupta, V.K. and Sorooshian, S. 1993. *A shuffled complex evolution approach for effective and efficient global minimization*. J. Optim. Theory Appl., **76**, 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**, 265-284.

Dunne, S.M., 1999. *Imposing constraints on parameter values of a conceptual hydrological model using baseflow response*. Hydrol. Earth Syst. Sci., **3(2)**, 271-284.

Evans J.P. and Jakeman, A.J. 1998. *Development of a simple, catchment-scale, rainfall-evapotranspiration-runoff model*. Env. Model. & Softw., **13**, 385-393.

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*. Wat. Resour. Res., **32(7)**, 2161-2173.

Gaume, E., Villeneuve, J.-P. and Desbordes, M. 1998. *Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model*. J. Hydrol., **210**, 38-50.

Grindley, J. 1970. *Estimation and mapping of evaporation. In: Symposium on world water balance (Proc. Reading Symposium), IAHS, Publ. no. 92, Vol. 1*, 200-213.

Gupta, V.K., and Sorooshian, S. 1985. *The relationship between data and the precision of parameter estimates of hydrologic models*. J. Hydrol., **81**, 57-77.

Gupta, H.V., Bastidas, L.A., Sorooshian, S., Shuttleworth, W.J. and Yang Z.L. 1998. *Parameter estimation of a land surface scheme using multicriteria methods*. J. Geoph. Res., **104(D16)**, 19491-19503.

Gupta, H.V., Sorooshian, S. and Yapo, P.O. 1998. *Towards improved calibration of hydrologic models: Multiple and noncommensurable measures of information.* Wat. Resour. Res., *34(4), 751-763.*

Hornberger, G.M. and Spear, R.C. 1981. *An approach to the preliminary analysis of environmental systems.* J. Env. Manage., *12, 7-18.*

Hornberger, G.M., Beven, K.J., Cosby, B.J. and Sappington, D.E. 1985. *Shenandoah watershed study: Calibration of the topography-based, variable contributing area hydrological model to a small forested catchment.* Wat. Resour. Res., *21(12), 1841-1850.*

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, 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 modeling the dynamic linkage of climate, physical catchment descriptors and hydrologic response components.* Mathematics and Computers in Simulation, *33, 359-366.*

Jakeman, A.J. and Hornberger, G.M. 1993. *How much complexity is warranted in a rainfall-runoff model?* Wat. Resour. Res., *29(8), 2637-2649.*

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, 297-314.*

Johnston, P.R. and Pilgrim, D.H. 1976. *Parameter optimization for watershed models.* Wat. Resour. Res., *12(3), 477-486.*

Jolley, T.J. 1995. *Large-scale hydrological modeling – The development and validation of improved land-surface parameterizations for meteorological input. Ph.D. Diss., Dep. Civil Env. Eng., Imperial College of Science, Technology and Medicine, London, UK, Unpublished.*

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.* Hydrol. Proc., *11, 1927-1944.*

Kirkby, M.J. 1975. *Hydrograph modeling strategies. In Peel, R., Chisholm, M., and Haggett, P. (eds.) Process in Physical and Human Geography, Heinemann, 69-90.*

Kleissen, F.M., Beck, M.B. and Wheater, H.S. 1990. *The identifiability of conceptual hydrochemical models.* Wat. Resour. Res., *26(12), 2979-2992.*

Kottegoda, N.T. and Rosso, R. 1997. *Statistics, Probability, and Reliability for Civil and Environmental Engineers. McGraw-Hill, USA.*

Kuczera, G. 1983. *Improved parameter inference in catchment models 1. Evaluating parameter uncertainty.* Wat. Resour. Res., *19(5), 1152-1162.*

Kuczera, G. and Mroczkowski, M. 1998. *Assessment of hydrologic parameter uncertainty and the worth of multiresponse data.* Wat. Resor. Res., *34(6), 1481-1489.*

Kuczera, G. and Parent, E. 1998. *Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm.* J. Hydrol., *211, 69-85.*

Leavesley, G. 1998. The modular modeling system (MMS) – A modeling framework for multidisciplinary research and operational applications. http: // wwwbrr.cr.usgs.gov / projects / SW_precip_runoff / mms /.

Lees, M.J. 2000. Data-based mechanistic modeling and forecasting of hydrological systems. J. Hydroinformatics, *02.1*, 15-34.

Lees, M.J., Price, N., Wheater, H.S. and Peach, D. 1998. A flood simulation model for the South Galway region of Ireland. In Proceedings of the BHS International Symposium on Hydrology in a Changing Environment, Vol. III, John Wiley & Sons, Chichester, UK, 93-104.

Lees, M.J., Wagener, T., Wheater, H.S., Gupta, H.V. and Beven, K.J. 1999. Assessing the regionalization potential of hydrological models. Proceedings of IUGG99, B.315, Birmingham, UK.

Lees, M.J. and Wagener, T. 2000a. The Monte Carlo Analysis Toolbox – User manual. Dep. Civil Env. Eng., Imperial College of Science, Technology and Medicine, unpublished.

Lees, M.J. and Wagener, T. 2000b. Review of HYCOM rainfall-runoff modeling undertaken for the review of Southern Water's Source Yields, 1997. Report to Southern Water, Imperial College Consultancy, London, UK.

Lees, M.J. and Wagener, T. 2000c. HYCOM rainfall-runoff modeling of Bewl-Darwell reservoir source rivers. Report to Southern Water, Imperial College Consultancy, London, UK.

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. Chapter 12, Institute of Hydrology Report No. 120, Wallingford, UK, 99-111.

Magee, B. 1973. Popper. William Collins Sons & Co, Glasgow.

Melching, C.S. 1995. Chapter 3: Reliability estimation. In Singh, V.P. (ed.) Computer models of watershed hydrology. Wat. Resour. Publ., USA, 69-118.

Mathworks 1996. Matlab – Reference guide. The Mathworks Inc., Natick, MA.

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, Berlin, 147-189.

Moore, R.J., and Clarke, R.T. 1981. A distribution function approach to rainfall-runoff modeling. Wat. Resour. Res., *17(5)*, 1367-1382.

Mous, S.L.J. 1993. Identification of the movement of water in unsaturated soils: the problem of identifiability of the model. J. Hydrol., *143*, 153-167.

Mroczkowski, M., Raper, G.P., and Kuczera, G. 1997. The quest for more powerful validation of conceptual catchment models. Wat. Resour. Res., *33(10)*, 2325-2335.

Naef, F. 1981. Can we model the rainfall-runoff process today? Hydrol. Sci. Bull., *26(3)*, 281-289.

Nash, J.E. and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models I: A discussion of principles. J. Hydrol., *10*, 282-290.

Niadas, I.A. 1999. *Rainfall-runoff model regionalization: An application to Midwest U.S.A.* M.Sc. Thesis, Dep. Civil Env. Eng., Imperial College of Science, Technology and Medicine, London, UK.

Overland, H., and Kleeberg, H.-B. 1993. *Moeglichkeiten der Abflussmodellierung unter Nutzung von Geoinformations-systemen.* Mitteilungen des Instituts fuer Wasserwesen, *45, Neubiberg, (in German).*

Penman, H.L. 1949. *The dependence of transpiration on weather and soil conditions.* J. Soil. Sci., *1, 74-89.*

Popper, K.R. 1983. *The logic of scientific discovery. 11th impression, Hutchinson, London, UK.*

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. 1992. *Numerical recipes in C: The art of scientific computing - 2nd Ed. Cambridge University Press, UK.*

Price, W.L. 1987. *Global optimization algorithms for a CAD workstation.* J. Optim. Theory Appl., *55(1), 133-146.*

Refsgaard, J.C. and Knudsen, J. 1996. *Operational validation and intercomparison of different types of hydrological models.* Wat. Resour. Res., *32(7), 2189-2202.*

Scholz, K., 1995. *Stochastic simulation of urban hydrological processes. (Stochastische Simulation urbanhydrologischer Prozesse). Ph.D. Diss, Fachbereich Bauingenieur- und Vermessungswesen, Universitaet Hannover, (in German), Germany, Unpublished.*

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, 1-16.*

Seibert, J. 1999. *Conceptual runoff models - fiction or presentation of reality? Acta Univ. Ups., Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology, 436, Uppsala.*

Sheratt, D.J. 1985. *Regional soil moisture modeling. Ph.D. Diss., Department of Civil Engineering, Imperial College of Science, Technology and Medicine, London, UK, Unpublished.*

Sorooshian, S. and Dracup, J.A. 1980. *Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases.* Wat. Resour. Res., *16(2), 430-442.*

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 model credibility.* Wat. Resour. Res., *19, 251-259.*

Spade, P.V. 2000. *Ockham's nominalist metaphysics: Some main themes. In: Spade, P.V. (ed.) The Cambridge companion to Ockham. Cambridge University Press, UK, 100-117.*

Spear, R.C. 1993. *Regional sensitivity analysis in environmental systems. In Young, P.C. (ed.) Concise encyclopedia of environmental systems. Pergamon Press, Oxford, UK, 476-479.*

Spear, R.C. 1995. *Large simulation models: Calibration, uniqueness and goodness of fit. MODSIM'95: International Congress on Modeling and Simulation, Model. Simul. Soc. Aust., Newcastle.*

Spear, R.C. and Hornberger, G.M. 1980. *Eutrophication in Peel Inlet, II, Identification of critical uncertainties via generalized sensitivity analysis,* Wat. Res., *14, 43-49.*

Spear, R.C., Grieb, T.M. and Shang, N. 1994. *Parameter uncertainty and interaction in complex environmental models.* Wat. Resour. Res., *30(11), 3159-3169.*

Sugawara, M. 1995. *Chapter 6: Tank Model. In Singh, V.P. (ed.) Computer models of watershed hydrology.* Wat. Resour. Publ., Colorado, USA, 165-214.

Uhlenbrock, S., Seibert, J., Leibundgut, C., and Rohde, A. 1999. *Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure.* Hydrol. Sci. J., *44(5), 779-797.*

Van Straten, G. and Keesman, K.J. 1991. *Uncertainty propagation and speculation in projective forecasts of environmental change: a lake-eutrophication example.* J. Forecasting, *10, 163-190.*

Wagener, T. 1998. *Developing a knowledge-based system to support rainfall-runoff modeling. MSc. Thesis, Dep. Civil Eng., Delft University of Technology, NL, Unpublished.*

Wagener, T., Lees, M.J. and Wheater, H.S. 1999. *A generic rainfall-runoff modeling toolbox. Eos Trans. AGU, 80(17), Fall Meet. Suppl., F203.*

Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V. and Sorooshian, S. 2000a. Submitted to* HESS.

Wagener, T., Lees, M.J., and Wheater, H.S. 2000b. *Incorporating predictive uncertainty into a rainfall-runoff modeling system. Accepted for Publication in the Proceedings of the Hydroinformatics Conference, Iowa, USA.*

Wang, Q.J. 1991. *The genetic algorithm and its application to calibrating conceptual rainfall-runoff models.* Wat. Resour. Res., *27(9), 2467-2471.*

Ward, R.C. and Robinson, M. 2000. *Principles of hydrology – 4th Ed..* McGraw-Hill, London.

Wheater, H.S., Bishop, K.H. and Beck, M.B. 1986. *The identification of conceptual hydrological models for surface water acidification.* Hydrol. Proc., *1, 89-109.*

Wheater, H.S., Jakeman, A.J. and Beven, K.J. 1993. *Chapter 5 – Progress and directions in rainfall-runoff modeling. In: Jakeman, A.J., Beck, M.B. and McAleer, M.J. (eds). Modeling change in environmental systems. John Wiley & Sons, UK, 101-132.*

Whitehead, P., Young, P.C. and Hornberger, G.M. 1979. *A systems model of stream flow and water quality in the Bedford-Ouse river – 1. Stream flow modeling.* Wat. Res., *13, 1155-1169.*

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.*

Wittenberg, H. 1999. *Baseflow recession and recharge as non-linear storage processes.* Hydrol. Proc., *13, 715-726.*

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, 25-30.

Yan, J. and Haan, C.T. 1991. *Multiobjective parameter estimation for hydrologic models – Weighting of errors.* Trans. Am. Soc. Agric. Eng*., **34(1)**, 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*., **18**, 23-48.*

Yapo, P.O., Gupta, H.V. and Sorooshian, S. 1998. *Multi-objective global optimization for hydrologic models.* J. Hydrol*., **204**, 83-97.*

Ye, W., Bates, B.C., Viney, N.R., Sivapalan, M., Jakeman, A.J. 1997. *Performance of conceptual rainfall-runoff models in low-yielding catchments.* Wat. Resour. Res*., **33**, 153-166.*

Young, P.C. 1985. *Recursive estimation and time-series analysis - An introduction.* Springer - Verlag, Berlin, Germany.

Young, P.C. 1992. *Parallel processes in hydrology and water quality: A unified time-series approach.* J. IWEM*, **6**, 598-612.*

Young, P.C., Parkinson, S. and Lees, M.J. 1996. *Simplicity out of complexity in environmental modeling: Occam's razor revisited.* J. Appl. Stat*., **23**, 165-210.*