HYDROLOGICAL PROCESSES
Hydrol. Process. **17**, 455–476 (2003)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.1135

Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis

T. Wagener,[1]* N. McIntyre,[1] M. J. Lees,[1†] H. S. Wheater[1] and H. V. Gupta[2]

[1] *Department of Civil and Environmental Engineering, Imperial College of Science, Technology and Medicine, Imperial College Road, London SW7 2BU, UK*

[2] *Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ, USA*

Abstract:

Conceptual modelling requires the identification of a suitable model structure and the estimation of parameter values through calibration against observed data. A lack of objective approaches to evaluate model structures and the inability of calibration procedures to distinguish between the suitability of different parameter sets are major sources of uncertainty in current modelling procedures. This paper presents an approach analysing the performance of the model in a dynamic fashion resulting in an improved use of available information. Model structures can be evaluated with respect to the failure of individual components, and periods of high information content for specific parameters can be identified. The procedure is termed dynamic identifiability analysis (DYNIA) and is applied to a model structure built from typical conceptual components. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS conceptual rainfall-runoff models; model structural analysis; parameter identifiability; information content of data

INTRODUCTION

Many, if not most, rainfall-runoff model structures currently used can be classified as conceptual. This classification is based on two criteria: (1) the structure of these models is specified prior to any modelling being undertaken; and (2) (at least some) of the model parameters do not have a direct physical interpretation, in the sense of being independently measurable, and have to be estimated through calibration against observed data (Wheater *et al.*, 1993). Calibration is a process of parameter adjustment (automatic or manual), until observed and calculated output time-series show a sufficiently high degree of similarity.

Conceptual rainfall-runoff (CRR) model structures commonly aggregate, in space and time, the hydrological processes occurring in a catchment into a number of key responses represented by storage components (state variables) and their interactions (fluxes). The model parameters describe aspects such as the size of those storage components, the location of outlets, the distribution of storages within the catchment, etc. Conceptual parameters, therefore, commonly refer to a collection of aggregated processes and they may cover a large number of subprocesses that cannot be represented separately or explicitly (Van Straten and Keesman, 1991). The usual underlying assumption, however, is that these parameters are, even if not measurable properties, at least constants and representative of inherent properties of the natural system (Bard, 1974: 11).

The modeller's task is the identification of an appropriate CRR model for a specific case, i.e. a given modelling objective, catchment characteristics, and data set (Wagener *et al.*, 2001a). A model is here defined as a selected model structure with a specific parameter set. Experience shows that this identification is a difficult task. Various parameter sets, often widely distributed within the feasible parameter space (e.g. Duan

* Correspondence to: T. Wagener, NSF Center for Sustainability of Semi-Arid Hydrology and Riparian Areas (SAHRA), Department of Hydrology and Water Resources, Harshbarger Building, PO Box 210011, Tucson, Arizona 85721-0011, USA.
E-mail: thorsten@sahra.arizona.edu

Received 31 January 2001
Accepted 2 January 2002

et al., 1992; Freer *et al*., 1996), and sometimes even different conceptualizations of the catchment system (e.g. Piñol *et al*., 1997; Uhlenbrook *et al*., 1999), may yield equally good results in terms of a predefined objective function. This ambiguity has serious impacts on parameter and predictive uncertainty (e.g. Beven and Binley, 1992), and, therefore, limits the applicability of CRR models, e.g. for the simulation of land use or climate-change scenarios, or for regionalization studies (Moore and Clarke, 1981).

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

Two reactions to this problem of ambiguity of system description can be found in the hydrological literature. The first is the increased use of parsimonious model structures (e.g. Jakeman and Hornberger, 1993; Young *et al*., 1996; Wagener *et al*., 2002), i.e. structures only containing those parameters, and therefore model components, that can be identified from the observed output. However, the increase in identifiability is at the price of a decrease in the number of separate processes described by the model. There is therefore a danger of building a model structure that is too simplistic for the anticipated purpose (Kuczera and Mroczkowski, 1998).

The second reaction is the search for calibration methods that make better use of the information contained in the available data time-series, e.g. streamflow and/or groundwater levels. Various research efforts have shown that the amount of information retrieved using a single objective function is sufficient to identify only between three and five parameters (e.g. Beven, 1989; Jakeman and Hornberger, 1993; Gupta, 2000). Most CRR model structures that use sub-monthly time steps contain a larger number. More information can become available through the definition of multiple objective functions to increase the discriminative power of the calibration procedure (e.g. Gupta *et al*., 1998; Gupta, 2000). These measures can either retrieve different types of information from a single time-series, e.g. streamflow (e.g. Gupta *et al*., 1998; Dunne, 1999; Boyle *et al*., 2000; Wagener *et al*., 2001a), or describe the performance of individual models with respect to different measured variables, e.g. groundwater levels (e.g. Kuczera and Mroczkowski, 1998; Seibert 2000) or saturated areas (Franks *et al*., 1998). Seibert and McDonnell (2001) show in a different approach how the parameter space can be constrained when soft data, i.e. qualitative knowledge of the catchment behaviour, is included in the calibration process. The soft data in their case included information, derived through experimental work, about the contribution of new water to runoff and also the restriction of parameter ranges to a desirable range based on experience. The result is a more realistic model, which will, however, yield sub-optimal performances with respect to many specific objective functions, in their case the Nash–Sutcliffe efficiency measure (Nash and Sutcliffe, 1970).

We therefore seek to increase the amount of information made available from an output time-series and to guide the identification of parsimonious model structures, consistent with a given model application as explained below. This paper presents a new approach to the identification and analysis of conceptual hydrological models called dynamic identifiability analysis (DYNIA) derived from the well-known regional sensitivity analysis (RSA; Spear and Hornberger, 1980; Hornberger and Spear, 1981). DYNIA is an attempt to avoid the loss of information through aggregation of the model residuals in time. This additional information can be used to analyse the working of the model, to find the amount of information available to identify a specific parameter, or to detect failures of underlying model assumptions in order to assess the adequacy of a selected model structure. The approach is applied to a conceptual model structure containing typical structural components and some initial results are presented.

IDENTIFIABILITY ANALYSIS OF CONCEPTUAL RAINFALL-RUNOFF MODELS

The purpose of identifiability analysis in CRR modelling is the identification of the model structure and a corresponding parameter set that are most representative of the catchment under investigation, while considering aspects such as modelling objectives and available data (Wagener *et al*., 2001a). This identifiability analysis can therefore be split into two stages: a model structure selection and a parameter estimation stage, which can, however, not be treated as completely separate (Sorooshian and Gupta, 1985).

Model structure identification

A large number of CRR modelling structures are currently available. These differ, for example, in the degree of detail described, the manner in which processes are conceptualized, requirements for input and output data, and possible spatial and temporal resolution. Despite these differences, a number of model structures may appear equally possible for a specific study, and the selection process usually amounts to a subjective decision by the modeller (Wagener, 1998), since objective decision criteria are often lacking (Mroczkowski *et al*., 1997). It is important, therefore, to deduce testable propositions with respect to the assumptions made in the model structure, i.e. about the hypothesis of how the catchment works, and to find measures of evaluation that give some objective guidance as to whether a selected structure is suitable or not. Uhlenbrook *et al*. (1999) have shown, however, that it is difficult to achieve this using single-objective Monte Carlo based calibration approaches. They were able to derive good performances from sensible and from incorrect conceptualizations of a catchment.

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

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

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

One, already mentioned, example of an underlying assumption of conceptual modelling is that the model parameters are usually considered to be constant in time, at least as long as, for example, no changes in the catchment occur that would alter the hydrological response (e.g. land-use changes). Different researchers (e.g. Beck, 1985; 1987; Gupta *et al*., 1998; Boyle *et al*., 2000; Wagener *et al*., 2001a) have shown that this assumption can be tested, and that the failure of a model structure to simulate different response modes with a single parameter set suggests inadequacies in the functioning of the model.

Beck used the extended Kalman filter (EKF) extensively to estimate model parameters recursively and to utilize the occurrence of parameter deviation as an indicator for model structural failure (e.g. Beck, 1985; 1987; Stigter *et al*., 1997). For example, in the identification of a model of organic waste degradation in a river, parameter value changes in time from one location in the parameter space to another were identified (Beck, 1985). Beck concluded from this variation that the model hypothesis had failed, i.e. the parameters

were changing to compensate for one or more missing aspect(s) in the model structure. The subsequent step is to draw an inference from the type of failure to develop an improved hypothesis of the model structure. However, there are limitations to the EKF approach. Beck (1985) concluded, with respect to the use of the EKF for hypothesis testing, that 'the performance of the extended Kalman filter (EKF) is not as robust as would be desirable and, *inter alia*, is heavily compromised by the need to make more or less arbitrary assumptions about the sources of uncertainty affecting the identification problem.'

Boyle *et al*. (2000) similarly showed for the example of a popular, moderately complex, rainfall-runoff model (Sacramento with 13 calibratable parameters) that a trade-off in the capability to simulate different response modes can occur; thus, it was not possible to reproduce (slow) recession periods and the remaining system response modes simultaneously. Their multi-objective analysis suggests that the cause for this problem was mainly an inadequate representation of the upper soil zone processes.

Parameter identification

The second stage in the model identification process is the estimation of a suitable parameter set, i.e. the actual calibration of the model structure. The parameters of each model structure are adjusted until the observed system output and the model output show acceptable levels of agreement. Manual calibration does this in a trial-and-error procedure, often using a number of different measures of performance and visual inspection of the hydrograph (e.g. Gupta *et al*., 1998). It can yield good results, but is time consuming, requires extensive experience with a specific model structure, and an objective analysis of parameter uncertainty is not possible. Traditional single-objective automatic calibration, on the other hand, is fast and objective, but will produce results that reflect the choice of objective function and may, therefore, often not be acceptable to hydrologists concerned with a number of aspects of performance (Boyle *et al*., 2000). In particular, the aggregation of the model residuals into an objective function leads to the neglect and loss of information about individual response modes, and can result in a biased performance, fitting a specific aspect of the hydrograph to the expense of another. It also leads to problems with the identification of those parameters associated with response modes that do not significantly influence the selected objective function (Wagener *et al*., 2001a). Selecting, for example, an objective function that puts more emphasis on fitting peak flows, e.g. the Nash–Sutcliffe efficiency value (Nash and Sutcliffe, 1970), due to its use of squared residual values (Legates and McCabe, 1999), will often not allow for the identification of parameters related to the slow response of a catchment (e.g. Dunne, 1999).

A comparison of hydrographs produced by different parameter sets that yield similar objective function values often shows that these hydrographs can be quite different. A 100 days extract of 6 years of daily streamflow data is shown in Figure 1, where the observed time series (black line) is plotted with seven different simulations (grey lines) using the same model structure, but different parameter sets. The objective function used during calibration is the root-mean-squared error (RMSE), which can be defined as follows

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(c_i - o_i)^2} \tag{1}$$

where c is the calculated flow at time step i, and o is the corresponding observed flow; N is the total number of time steps considered. All models yield an RMSE of 0·60 mmd^{-1} when the complete calibration period (6 years) is considered. However, the hydrographs produced are clearly visually different. This demonstrates that traditional single-objective optimization methods do not have the ability to distinguish between visually different behaviour (Gupta, 2000). The requirement for a parameter set to be uniquely locatable within the parameter space, i.e. to be globally identifiable, is that it yields a unique response vector (Kleissen *et al*., 1990; Mous, 1993). The unique response vector, in this case a unique (calculated) hydrograph, might be achievable, but this uniqueness is lost if described by a single objective function. Such problems cannot be solved through improved search algorithms. They are rather inherent in the philosophy of the calibration procedures.

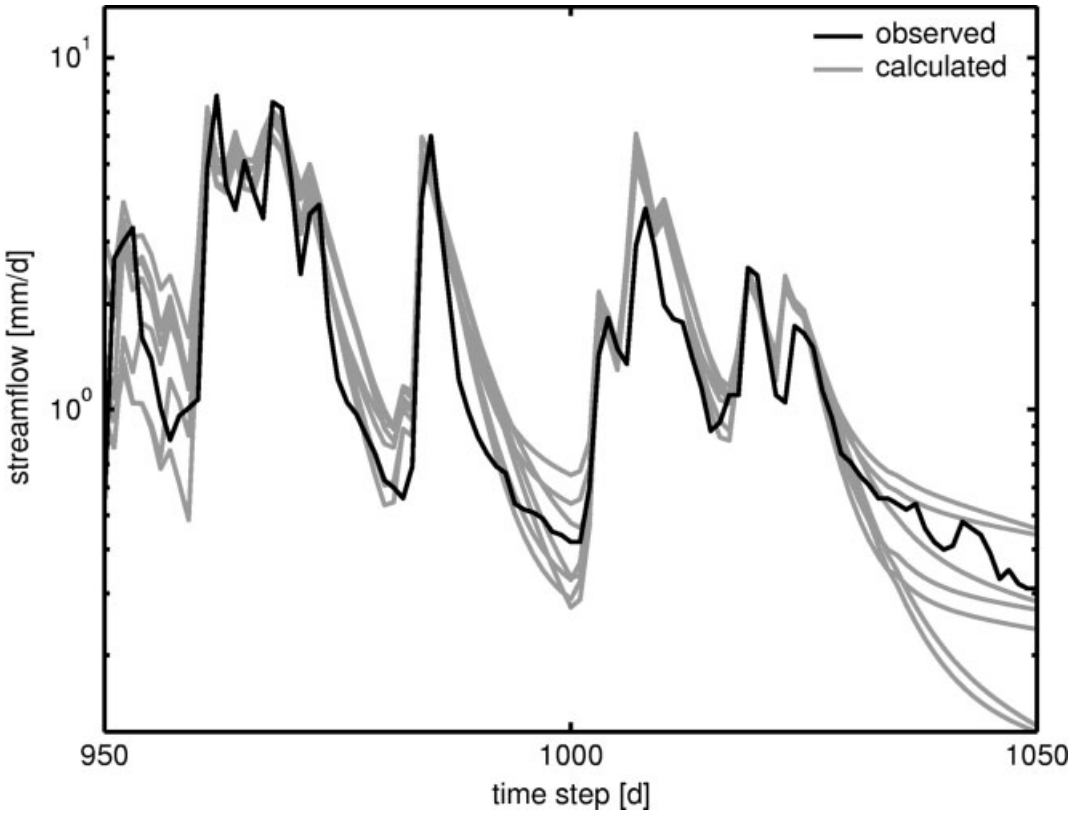

Figure 1. An extract of a 6 year calibration period using daily data. The observed streamflow is shown in black, and seven different model realizations are plotted in grey. All simulations yield an RMSE value of 0·60. The streamflow is plotted on a logarithmic scale

Clearly, the complex thought processes, which lead to successful manual calibration, are very difficult to encapsulate in a single objective function. This is illustrated by the requirements defined by the US National Weather Service (NWS) for the manual calibration of the Sacramento model structure (NWS, 2000):

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

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

Gupta *et al.* (1998) review this problem in more detail and conclude that a multi-objective approach to automatic calibration can be successful. Boyle *et al.* (2000) show how such a procedure can be applied to combine the requirements of manual calibration with the advantages of automatic calibration. A multi-objective algorithm is used to find the model population necessary to fit all aspects of the hydrograph. The user can then, if necessary, manually select a parameter set from this population to fit the hydrograph in the desired way. This will, however, in the presence of model structural inadequacies, lead to a sub-optimal performance with respect to at least some of the other measures (Boyle *et al.*, 2000; Seibert and McDonnell, 2001). The resulting trade-off of the ability of different parameter sets to fit different aspects of the hydrograph usually leads to a compromise solution (Bastidas, 1998; Ehrgott, 2000). Their procedure analyses the local behaviour of the model in addition to its global behaviour (Gupta, 2000). The global behaviour is described through objective functions, such as overall bias or some measure of the overall variance, e.g. the RMSE. The local behaviour is defined by aspects like the timing of the peaks, or the performance during quick and slow response periods (Boyle *et al.*, 2000; 2001).

Multi-objective calibration of single-output models

One way of implementing automatic multi-objective calibration is by partitioning the continuous output time series into different response periods. A separate objective function can then be specified for each period, thus reducing the amount of information lost through aggregation of the residuals.

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

Though these methods help to retrieve more information, they also show some weaknesses. Approaches (a) and (b) are subjective and based on the hydrologist's experience, and so are not easily applicable to a wide variety of models and catchments. Approach (c), while objective, does not recognize the effects of parameter dependencies, and may not highlight periods that are most informative about the parameters as independent entities. The sensitivity of the model performance to changes in the parameter is a necessary requirement, but it is not sufficient for the identifiability of the parameter. Furthermore, if the parameter sensitivity is measured locally (e.g. Kleissen, 1990), the result is not guaranteed over the feasible parameter space. However, Wagner and Harvey (1997) show that some of these problems can be reduced by implementing a Monte Carlo procedure where the sensitivity for a large number of different parameter combinations is assessed using parameter covariance matrices. Approach (d) is independent of any model structure and links between the results the model parameters still need to be established.

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

DYNIA

DYNIA is a new approach for locating periods of high identifiably for individual parameters and to detect failures of model structures in an objective manner. The proposed methodology draws on elements of the popular RSA (Spear and Hornberger, 1980; Hornberger and Spear, 1981), the Generalized Likelihood Uncertainty Estimation Framework (GLUE, Beven and Binley, 1992), wavelet analysis (Goswami and Chan, 1999), and applications of the EKF for hypothesis testing by Beck (1985, 1987).

The basis of the original RSA approach is an investigation of whether the parameter distribution changes when it is conditioned on a measure of performance, e.g. an objective function. Deviations from an initially uniform distribution, and differences between those parts of the distribution performing well and poorly, often called behavioural and non-behavioural, indicate the sensitivity of the model response to changes in the parameter (Spear and Hornberger, 1980). The approach is extended here to assess the identifiability of parameters, not just their sensitivity.

The basic steps in the procedure can be seen in the flow chart in Figure 2. Monte Carlo sampling based on a uniform prior distribution is used to examine the feasible parameter space. The objective function associated with each parameter set is transformed into a support measure, i.e. all support measures have the characteristic that they sum to unity, and higher values indicate better performing parameter values. These are shown here in form of a dotty plot (Figure 2a). The best performing parameter values (e.g. top 10%) are selected and their cumulative support is calculated (Figure 2b). A straight line will indicate a poorly identified parameter, i.e. the highest support values are widely distributed over the feasible range. Deviations from this straight line indicate that the parameter is conditioned by the objective function used. The gradient of the cumulative support is the marginal probability distribution of the parameter, and therefore an indicator of the strength of the conditioning, and of the identifiability of the parameter. Segmenting the range of each parameter (e.g. into 20 containers) and calculating the gradient in each container leads to the (schematic) distribution in Figure 2d. The highest value, additionally indicated by the darkest colour, marks the location (within the chosen resolution) of greatest identifiability of the parameter. Wagener *et al*. (2001a) show how this measure of identifiability can be used to compare different model structures in terms of parameter uncertainty, which is assumed to be the opposite of identifiability. They calculate the identifiability as a function of measures of performance for the whole calibration period and for specific response modes, derived using the segmentation approach by Boyle *et al*. (2000) described earlier in the text.

Calculating the parameter identifiability at every time step by using the residuals for a number of time steps n before and after the point considered, i.e. a moving window or running mean approach, allows the investigation of the identifiability as a function of time (Figure 2e). The gradient distribution plotted at time t, therefore, aggregates the residuals between $t - n$ and $t + n$, with the window size being $2n + 1$. The number of time steps considered depends upon the length of the period over which the parameter is influential. For example, investigation of a slow response linear store residence time parameter requires a wider moving window than the analysis of a quick response residence time parameter. Window sizes (of 11, 21, 41, and 101 time steps) are used in the example application presented later in the text. A diversity of model structures will be tested in the future to provide guidance on appropriate window sizes. A window that is too small can be greatly influenced by data error. However, small window sizes can be used in cases where the data quality is very high, for example in the case of tracer experiments in rivers (Wagener *et al*., 2001b). Conversely, if the window size is too big, periods of noise and periods of information will be mixed and the information will be blurred.

The results are plotted for each parameter versus time using a colour coding, where a darker colour indicates areas, in parameter space and time, of higher identifiability. Care has to be taken when interpreting the DYNIA

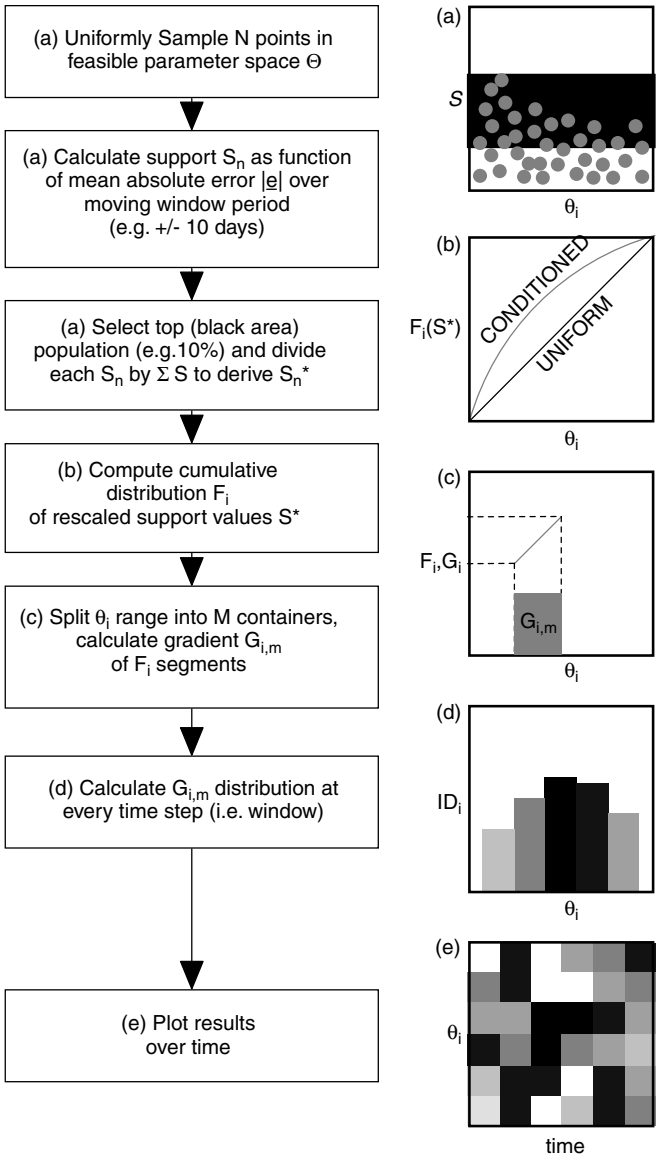

Figure 2. The DYNIA procedure

results of time steps at the beginning and the end of time series. Here, the full window size cannot be used and the result is distorted. This is an effect similar to the cone of influence in wavelet analysis (Torrence and Compo, 1998).

Although this approach is not intended to evaluate parameter dependencies in detail, the significance of dependencies to the identifiability is implicit in the univariate marginal distribution that is structurally represented by Figure 2d. A strong dependency during any period would tend to inhibit the information of a strong univariate peak, i.e. the effect of the parameters involved cannot be singled out. Parameter interdependence can be estimated in detail by the investigation of the response surface or the variance–covariance matrix (e.g. Hornberger *et al*., 1985; Wheater *et al*., 1986).

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

Using parameter variation as an indicator of model structural failures assumes, of course, that the specific parameter does not describe characteristics of the catchment that are time variant, for example the leaf area. Variation in good parameter values in those cases would rather corroborate the model structure, and not indicate a failure.

APPLICATION EXAMPLE

Data and model structure

The river selected for this study is the Lower Medway at Teston (1256·1 km^2) located in southeastern England. Six years (10 April 1990–14 July 1996) of data (daily naturalized flows, precipitation, potential evapotranspiration (PE), and temperature) are available (Figure 3). The Medway catchment is characterized

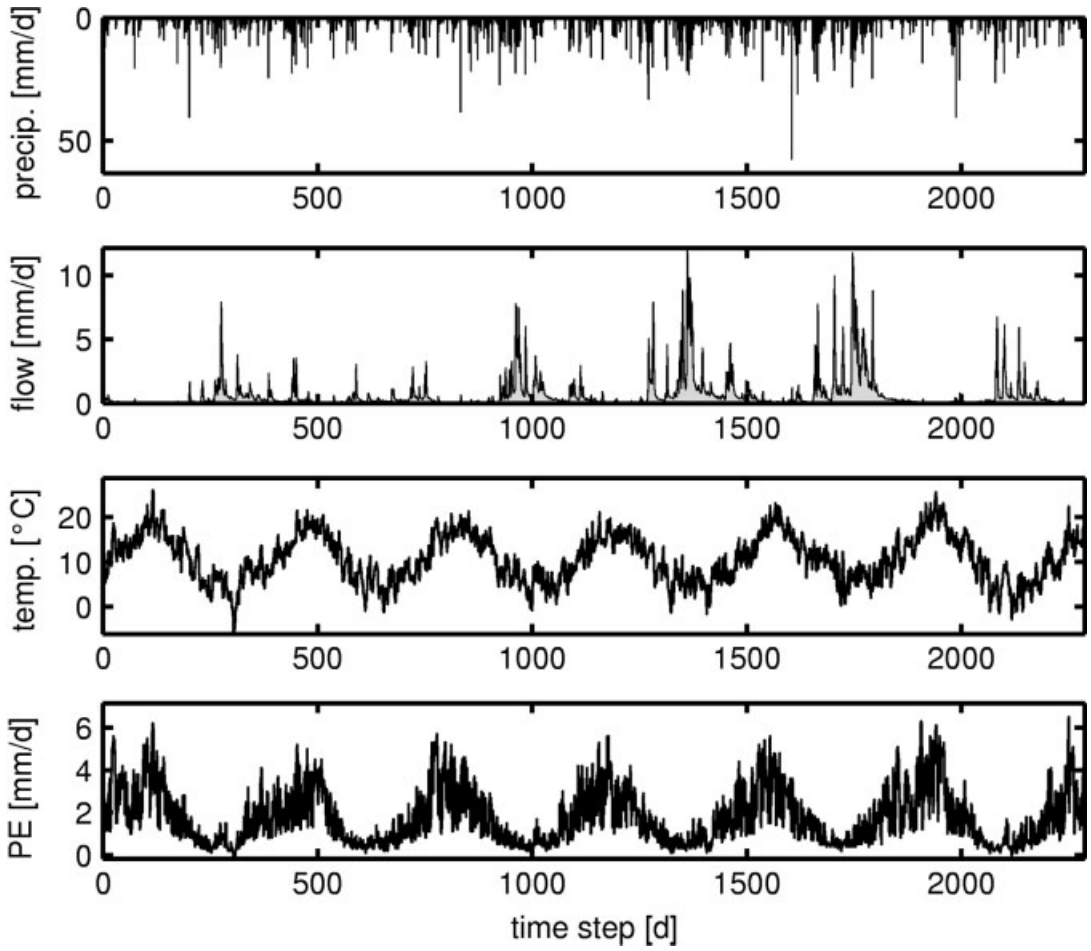

Figure 3. Time series used in the application example. Six years (10 April 1990–14 July 1996) of data (daily naturalized flows, precipitation, PE, and temperature) for the Lower Medway at Teston

by a mixture of permeable (chalk) and impermeable (clay) geologies subject to a temperate climate with an average annual rainfall of 772 mm and an average annual PE of 663 mm (1990–96).

The Rainfall-Runoff Modelling Toolbox (RRMT) and Monte Carlo Analysis Toolbox (MCAT), developed at Imperial College, are used for calculation and visualization of results (Wagener *et al*., 1999, 2002). The RRMT is a generic modelling shell allowing its user to implement different lumped model structures of conceptual or hybrid metric–conceptual type. Hybrid metric–conceptual models utilize observations to test hypotheses about the model structure at the catchment scale and, therefore, combine the metric and the conceptual paradigm (Wheater *et al*., 1993). The structure selected here is a combination of a Penman-type soil water accounting component (Penman, 1949), as used by Jolley (1995), and a parallel routing structure consisting of two linear conceptual reservoirs to represent quick and slow catchment response (Figure 4). The ratio of flow contributing to each of the two reservoirs is fixed. The model structure contains five parameters in total. The Penman model has two parameters, the size of the near-surface store, defined by a root constant 'rc', plus an additional 25 mm to allow for capillary rise (Penman, 1949), and a 'bypass' parameter. The bypass component represents phenomena that divert water from the soil water store and lead to rapid groundwater recharge or runoff response during rainfall, such as macropore and infiltration excess overland flow (Jolley, 1995). It applies to the proportion of the rainfall that exceeds the PE. The near-surface store is connected by an overflow mechanism to the lower store. The size of the lower store is chosen large enough to ensure that it never empties (Moore, 1999). Additional effective rainfall is produced when both stores are filled and the lower store overflows. Evapotranspiration takes place at the potential rate from the near-surface store. It reduces to 1/12 of the potential rate from the lower store when the upper store is emptied, as suggested by Penman (1949). The split of the effective rainfall between quick and slow flow is defined by the parameter α, which is the ratio of flow going towards the quick response reservoir. The remaining two parameters are the residence times of the two linear stores rt(q) and rt(s).

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

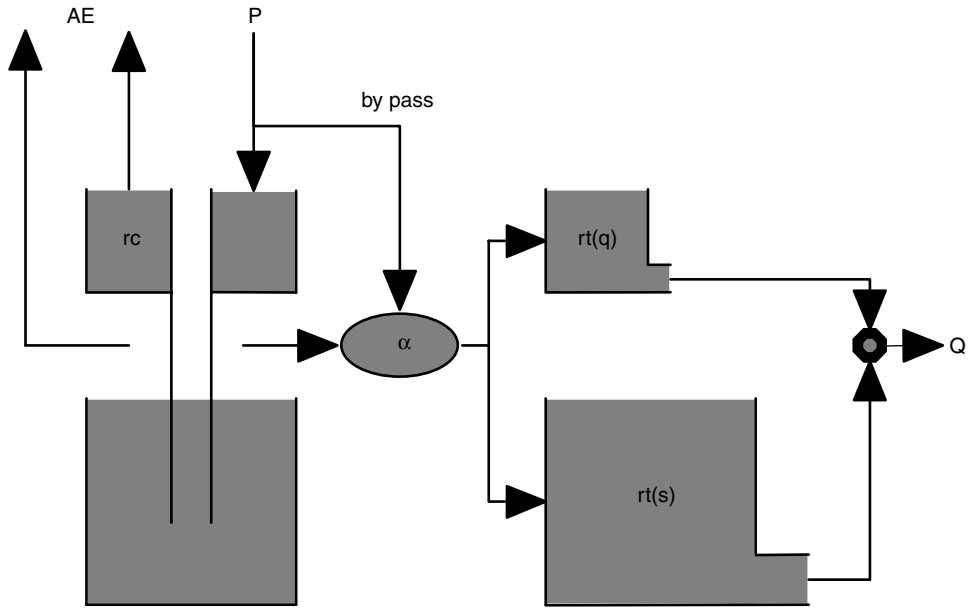

Figure 4. The model structure used in the application example

RESULTS AND DISCUSSION

Traditional Monte Carlo sampling

The result of a conventional Monte Carlo procedure based on a uniform distribution, sampling 20 000 points in the feasible parameter space, is shown in the form of dotty plots in Figure 5. This is used as a benchmark for evaluation of the DYNIA results. The objective function used in Figure 5 is the RMSE defined earlier in the text, in this case using the residuals over the whole 6 year period. It can be seen from these plots that some of the parameters show quite a distinct optimum, e.g. rt(q) or α, whereas others (in particular rt(s)) reveal equally performing values over a relatively widespread part of the feasible parameter space.

However, some of the response surface structure in parameter space (Beven, 1998) and in time, as will be shown later, is lost through the projection into a single dimension in these dotty plots.

Information content

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

The information content is calculated using the first two steps of the DYNIA analysis, shown in Figure 2a and b. The cumulative distributions calculated for every time step (Figure 2b) can be used to derive confidence limits for the different parameters (e.g. 90%). Wide confidence limits suggest that parameter values associated with equally good performance are distributed widely over the parameter space; narrow limits suggest that the best performing parameters are concentrated in a small area of the feasible range. A transformed measure (one minus the width of the confidence limits over the parameter range, which is normalized to run from zero to one) is used here so that a large value is equal to a high information content for a given time step. The time series of the information measure is plotted for each of the parameters in Figure 6, together with

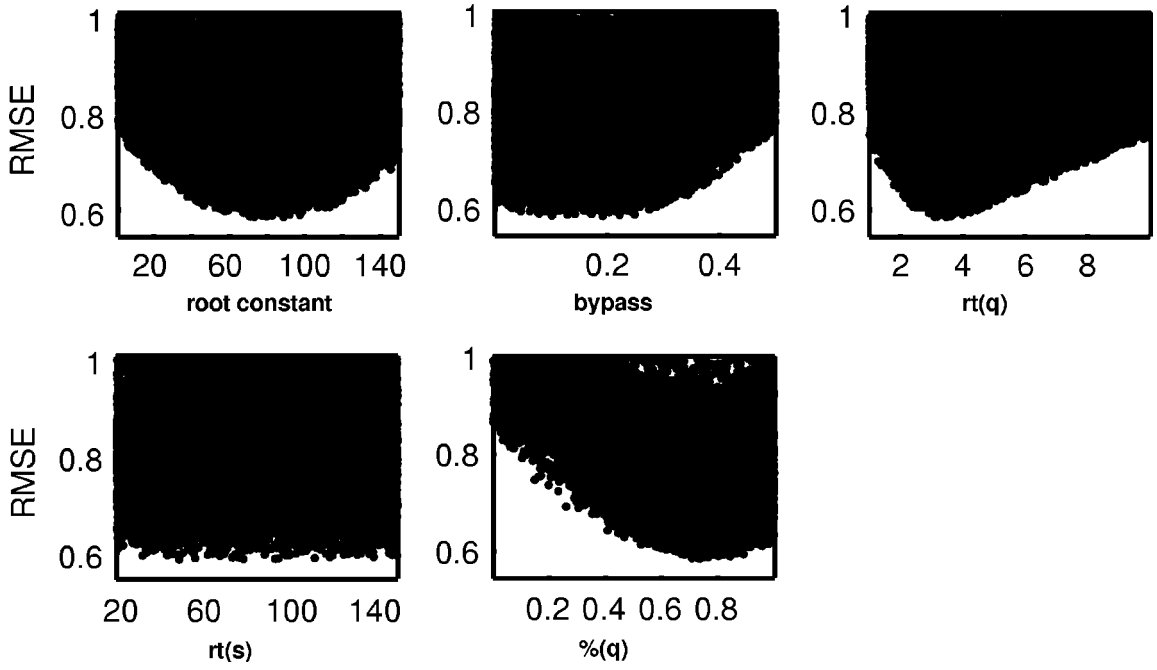

Figure 5. Dotty plot showing results of the uniform random search using 20 000 samples utilising the whole 6 year period available. Lower values of the RMSE objective function indicate better performing parameter values. Only parameter sets producing an RMSE below one are shown

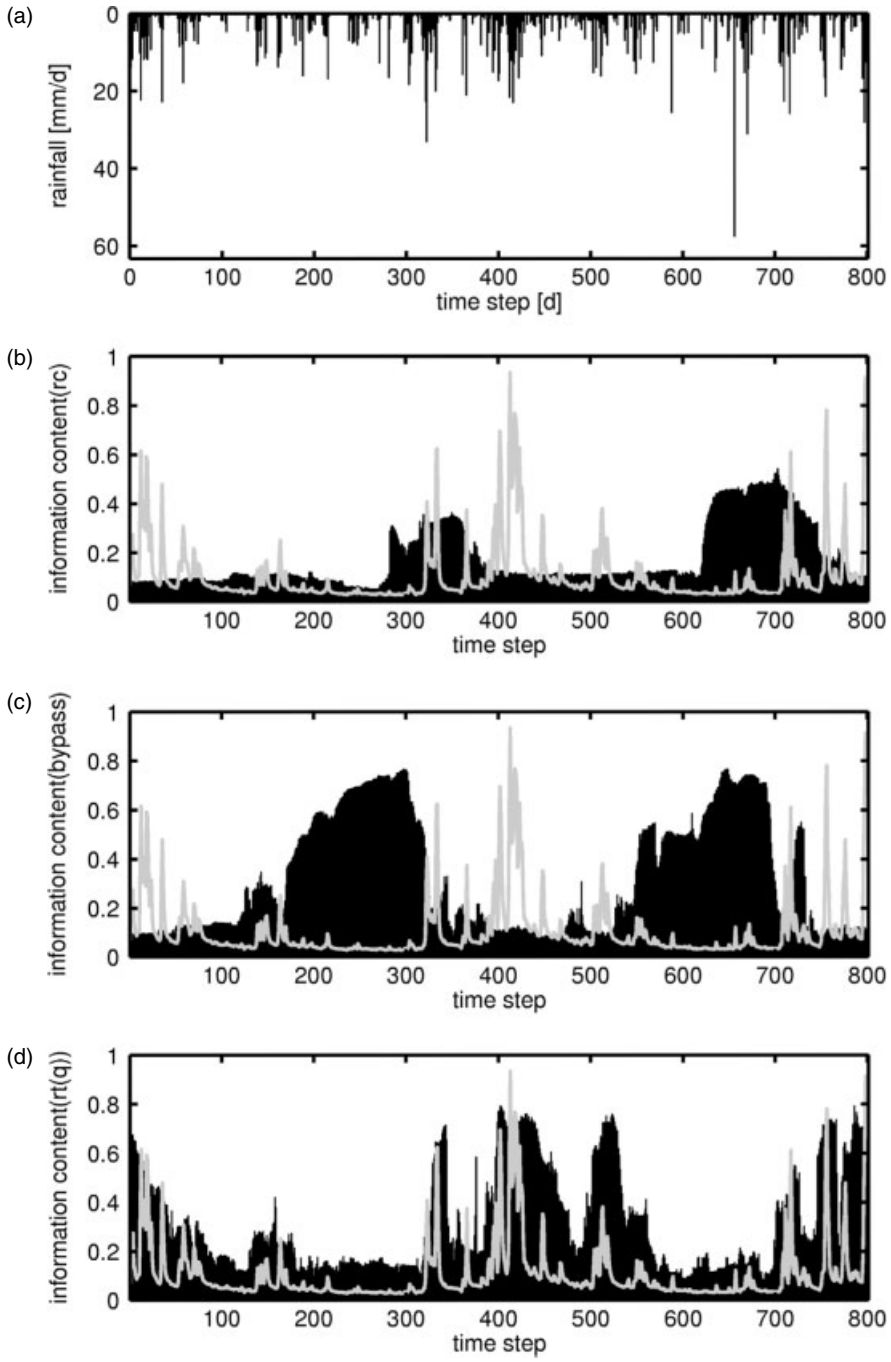

Figure 6. The information content of the data over a 2 year period (days 950 to 1750). Graph (a) shows the precipitation input over the period considered. The remaining plots show the result for the different parameters (black bars). The grey line is the streamflow, normalized with respect to its maximum value

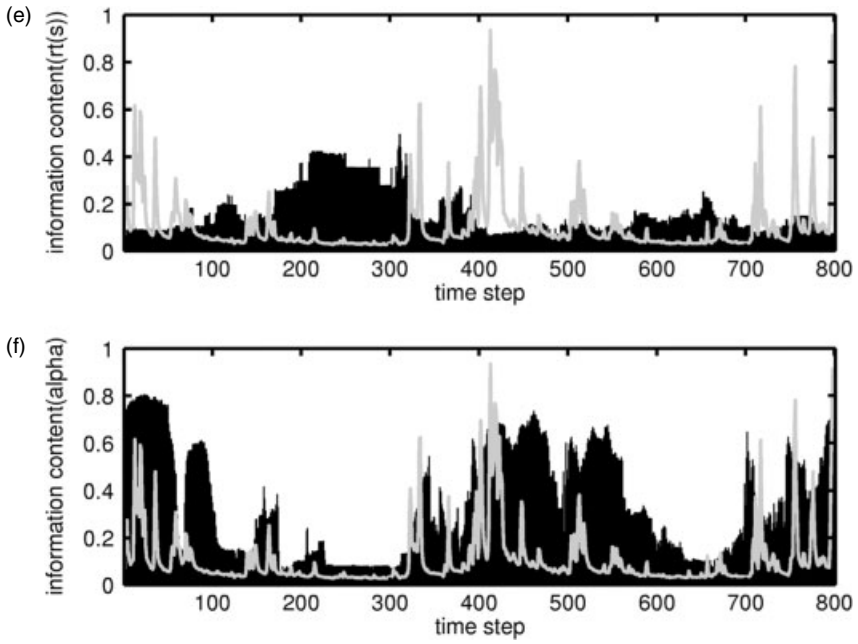

Figure 6. (*Continued*)

the streamflow (normalized for display) and the rainfall. It is important to remember that this plot contains a subjective element through the definition of the initial feasible parameter space by the modeller.

Figure 6b shows the information content for the root constant 'rc' derived using a window size of 101 (daily) time steps. It can clearly be seen that the main information about the root constant emerges towards the end of long recession periods (dry summers) and, in particular, during the wetting up periods. The information values during the remaining periods are relatively small.

The 'bypass' parameter is analysed in Figure 6c with a window of 41 time steps. The plot reveals three types of period where information is available. The first is for small runoff events after wet winter periods, e.g. around days 150 and 500, the second is located at summer periods, e.g. days 175–300 and 550–700, and the third during wetting-up phases, e.g. days 350 and 725.

Information about the quick flow residence time rt(q) (Figure 6d, using a window of 11 time steps) can mainly be found during the quick recessions after high flow events, whereas the long recession tails contain the information about the slow flow residence time rt(s) (Figure 6e, using a window of 41 time steps). Using larger window sizes for rt(s) did not improve the result. An attempt at using a regressive variant of the moving window approach, in which only a certain number of time steps up to the time step itself are considered, to improve the results for the residence time parameters, was not successful. This is because, especially when large window sizes are used, periods of high identifiability are shown after the time steps that actually contain the information.

The analysis of the split parameter α (Figure 6f, using a window size of 21 time steps) reveals that this parameter becomes identifiable after flow events when the response is changing from quick to slow drainage. Little information about this parameter can be gathered during periods of long recessions with only minor runoff events.

The information contained in Figure 6 can also be used to find combinations of parameters responsible for the model's behaviour during specific response periods. These interacting parameters could then be grouped for multivariate calibration (e.g. Wheater *et al.*, 1986). A threshold for the information content value of 0·3 is, somewhat arbitrarily, selected here, and the selected high information content time steps for the different

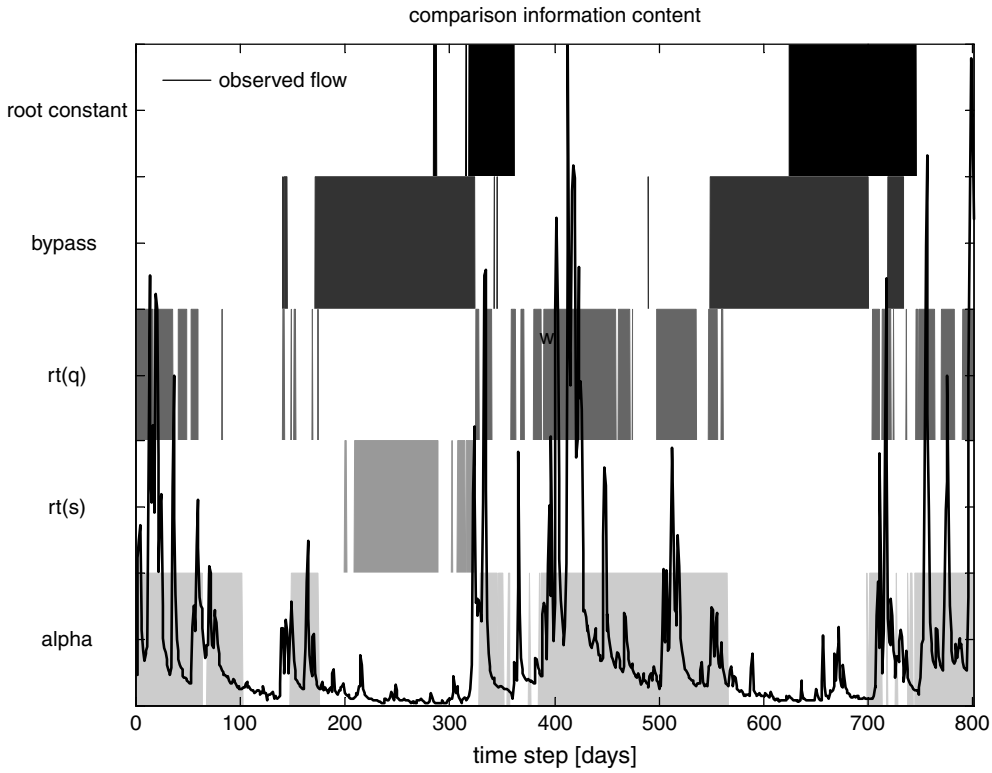

Figure 7. Comparison of the information content of a 2 year period for the different parameter values (days 950 to 1750). Only time steps with an information content above 0·3, with respect to individual parameters, are shown in the different grey shades

parameters are shown in Figure 7 (each parameter is indicated by a different grey shading) together with the normalized streamflow. From this plot it is easy to see that rt(q) and α show high information contents during similar periods and, therefore, during the same response modes. There is also a considerable overlap between the relevant periods for the 'bypass' and rt(s) parameters, at least during the first slow recession phase. However, these similarities do not necessarily imply parameter interdependence.

The initial Monte Carlo simulation over the whole calibration period was based on the RMSE. It is clear from Figure 5 that this measure is not capable of retrieving information to distinguish between the performance of different values of rt(s), which only becomes identifiable during distinctive periods of recession. The RMSE emphasizes performance during peak flow periods. However, applying a simple threshold to the data, separating out periods of low flow, can improve identifiability, as demonstrated in Figure 8. The three different lines display the gradient distributions over the range of rt(s). The gradients are derived using steps (a) and (b) of the DYNIA procedure (Figure 2), where the feasible range is split into ten containers of equal size. These gradients represent the full data record, and those time steps where the observed flow is below a certain threshold, i.e. 0·5 mm day^{-1}, in order to consider only periods of recession. It can be seen that the identifiability is improved. The dashed line shows the result when only time steps below the selected threshold and with an information content of above 0·15 are considered. The additional flow criterion is required, since informative regions can also be found during high flows showing different optima (lower values) in the parameter space. This is caused by structural inadequacies in the simple slow flow component of the model. The dashed line shows the highest identifiability values and reveals that a lower value of rt(s) performs better when the influence of high flow periods is removed.

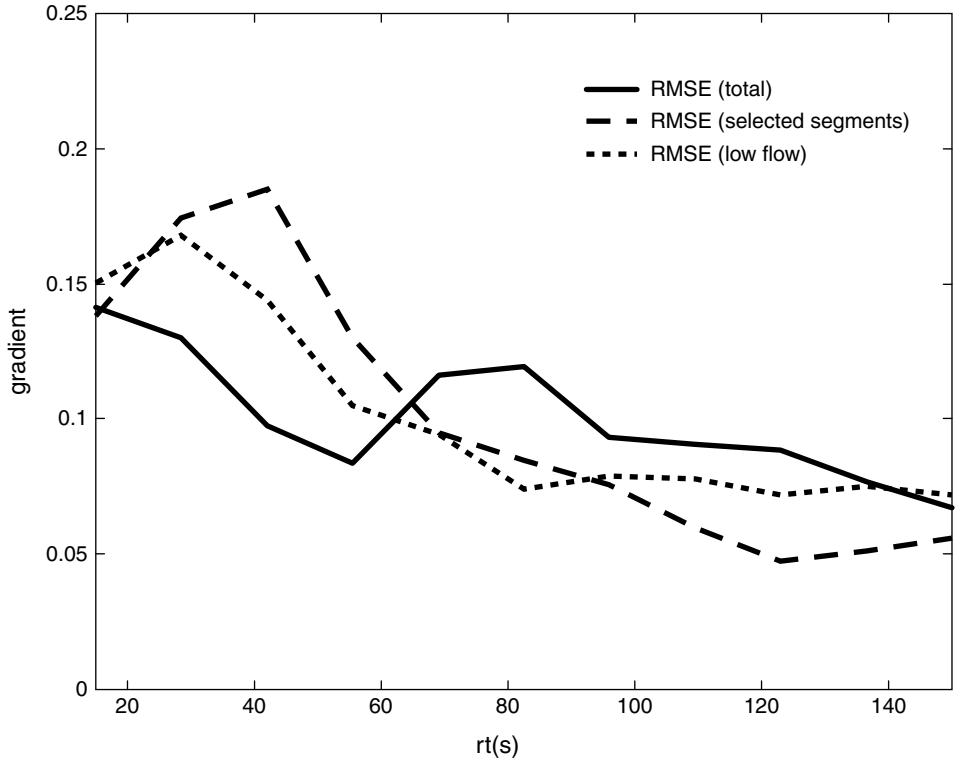

Figure 8. Comparison of the identifiability, defined as the gradients of the cumulative distribution at different locations of the parameter range, of the slow response residence time rt(s) using the RMSE, (1) as an overall measure of the performance over the whole calibration period (continuous line), (2) only utilizing the residuals at time steps with flow values below 0·5 mm day^{-1} (dotted line), and (3) using residuals with flow values below 0·5 mm day^{-1} and with an information content above 0·15 (dashed line)

Dynamic identifiability and structural failure

The information content plots only describe where in the time series a parameter becomes identifiable. They do not give any information about the location of optima in the parameter space. A different type of plot is therefore shown in Figure 9, derived by performing the remaining stages of DYNIA. The plots visualize the DYNIA results in the parameter–time space. The values of the identifiability measure are transformed into a grey shading, with higher values indicated by a darker colour, and plotted against the time axis (see Figure 2c and d). Additionally, the 90% confidence limits, derived from the cumulative distributions, and the streamflow, normalized with respect to its maximum value, are shown.

Figure 9b shows the results for the root constant 'rc'. It can be seen that the confidence limits narrow during the wetting-up periods after the dry summers. During those periods, the parameter clearly strives to higher values. No particular optima are visible during the remaining periods, indicating that very different values of this parameter yield similar results in combination with the remaining parameters. There are, however, two different optima visible in Figure 9b: one around a value of 80 and the other around 120. These different optima could be related to the problem of simultaneously fitting the overall water balance and the wetting-up periods with a single parameter. This parameter requires the largest window for its analysis. Investigations, so far, suggest that this is typical for parameters describing the maximum storage capacity.

The 'bypass' parameter has different periods containing high information content, as shown in Figure 6c. In Figure 9c, one can see that the parameter jumps between at least two optima. Small values of this parameter perform better during low flow periods, e.g. around time steps 250–300 and 650–700. During other periods,

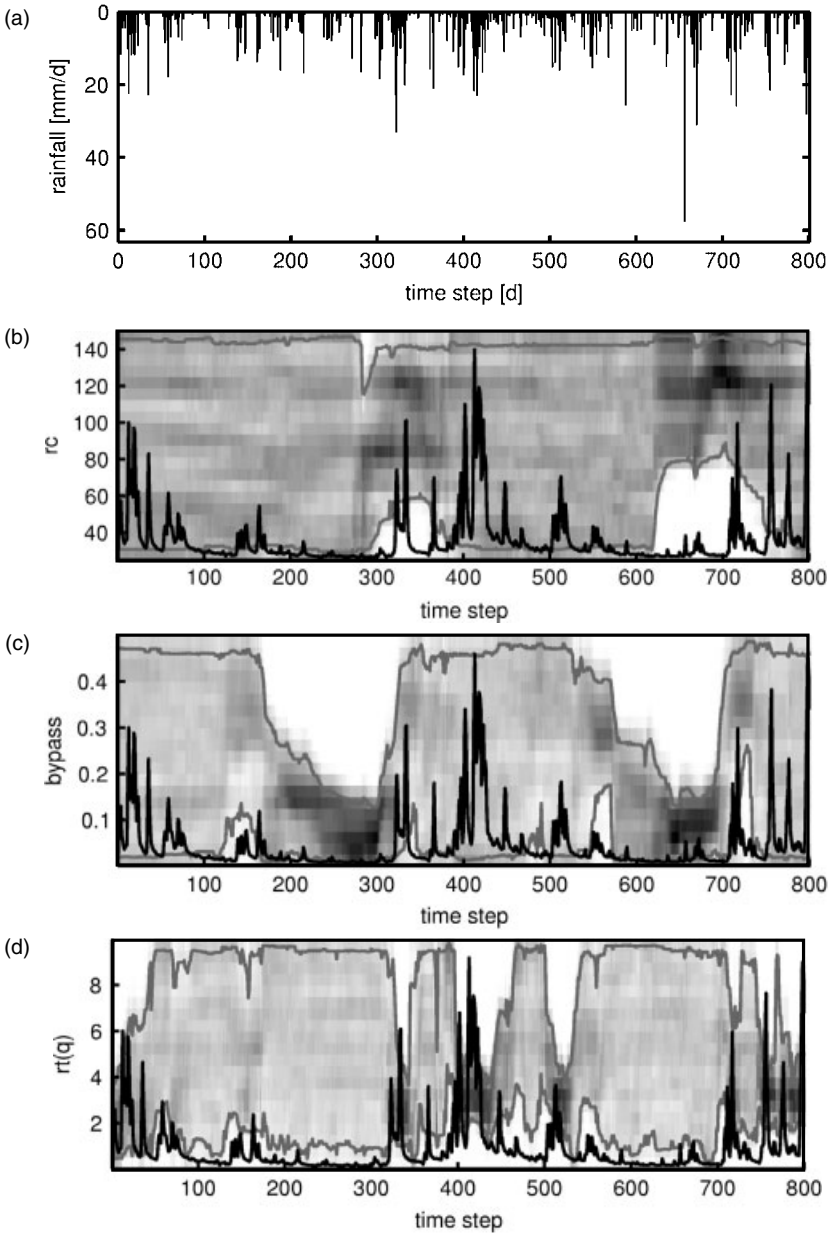

Figure 9. Results of the DYNIA procedure for a 2 year period (days 950 to 1750). Graph (a) shows the rainfall input over this time. The remaining graphs show the DYNIA results for the different parameters. The grey shading indicates the size of the gradient, with a darker colour for a higher value. The dark grey lines are the 90% confidence limits derived from the cumulative distribution of support values, and the black line is the streamflow normalized with respect to its maximum value

e.g. around time steps 150, 550, and 725, larger values of the 'bypass' (0·3 or higher) seem to provide a better fit to the observed data, albeit with decreased identifiability (lighter grey shading). Small 'bypass' values are required during summer periods to yield only little runoff during summer storms when the catchment is dry. High values are needed during storm events after the wet periods and in the wetting up periods. This can explain why the Monte Carlo simulation results shown in Figure 5, using the RMSE, provide no clear

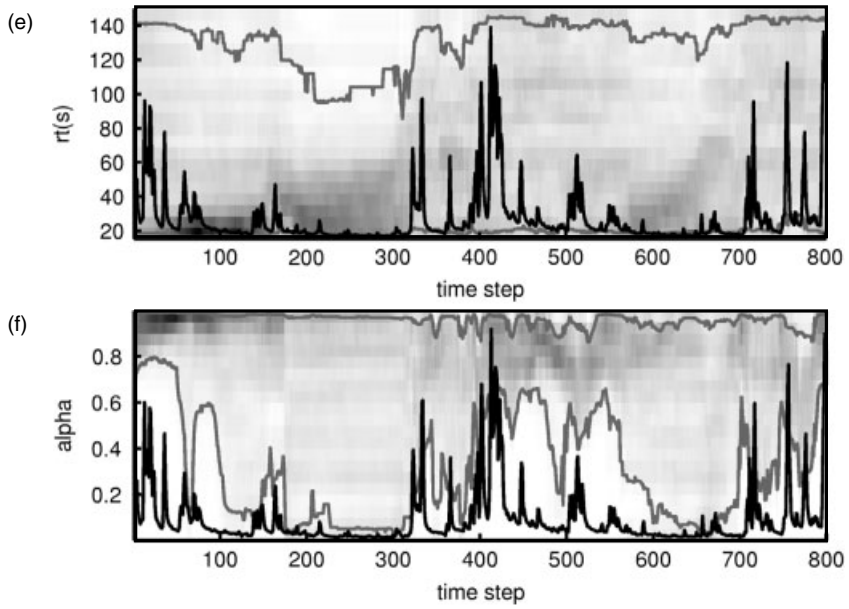

Figure 9. (*Continued*)

optimum with respect to this parameter. The RMSE measure is biased towards fitting higher flow values, and the high identifiability areas during slow recessions might not be influential enough to produce a clear peak on the dotty plots. This change in optimum parameter value within the parameter space is an indicator of failure of the model structure. There is a clear inconsistency in the way the model fits the observed behaviour of the catchment.

Parameter rt(q) is analysed in Figure 9d. Clear optima occur at a value of approximately three during the quick recession periods, whereas there are no specific peaks during other time steps.

It is, at least, very difficult to identify suitable values for rt(s) using a measure like RMSE. As noted above, the residuals of slow flow periods are often too small to influence the overall performance of a model, as shown in Figure 5. Figure 9e, however, shows that better-performing values for this parameter lie near to the lower boundary of the feasible parameter space, especially at time steps 200–300, which is the longest recession period contained in the data set used.

Parameter α does not show such a distinct area of identifiability, such as for example 'rc' or 'bypass', apart from a short initial period within the warm-up range of the algorithm (Figure 6f). There is, however, some area of darker grey shading in the period between time steps 400 and 550, which is after the main wet period. The parameter varies roughly between 0·6 and 0·9 during this period, which is also the range found in the initial Monte Carlo analysis (Figure 5). Figure 6f suggests some variation of this parameter in time. However, the evidence here is not sufficient to draw conclusions, and further analysis is required.

Two-dimensional response surface plots are used in Figure 10 to analyse the identifiability and interaction between the soil moisture accounting parameters, i.e. the root constant and the bypass, more closely. The upper plot shows the result when the RMSE is calculated using the residuals for the complete time series, and the bottom plot only uses time steps with high information content and excludes those that show ambiguity with respect to the bypass parameter. One can see that the parameters are much better identified when periods of noise are not considered. The optimum values for the root constant are, however, slightly smaller than suggested in the DYNIA plots. This emphasizes the need for a detailed analysis using different methods of visualization.

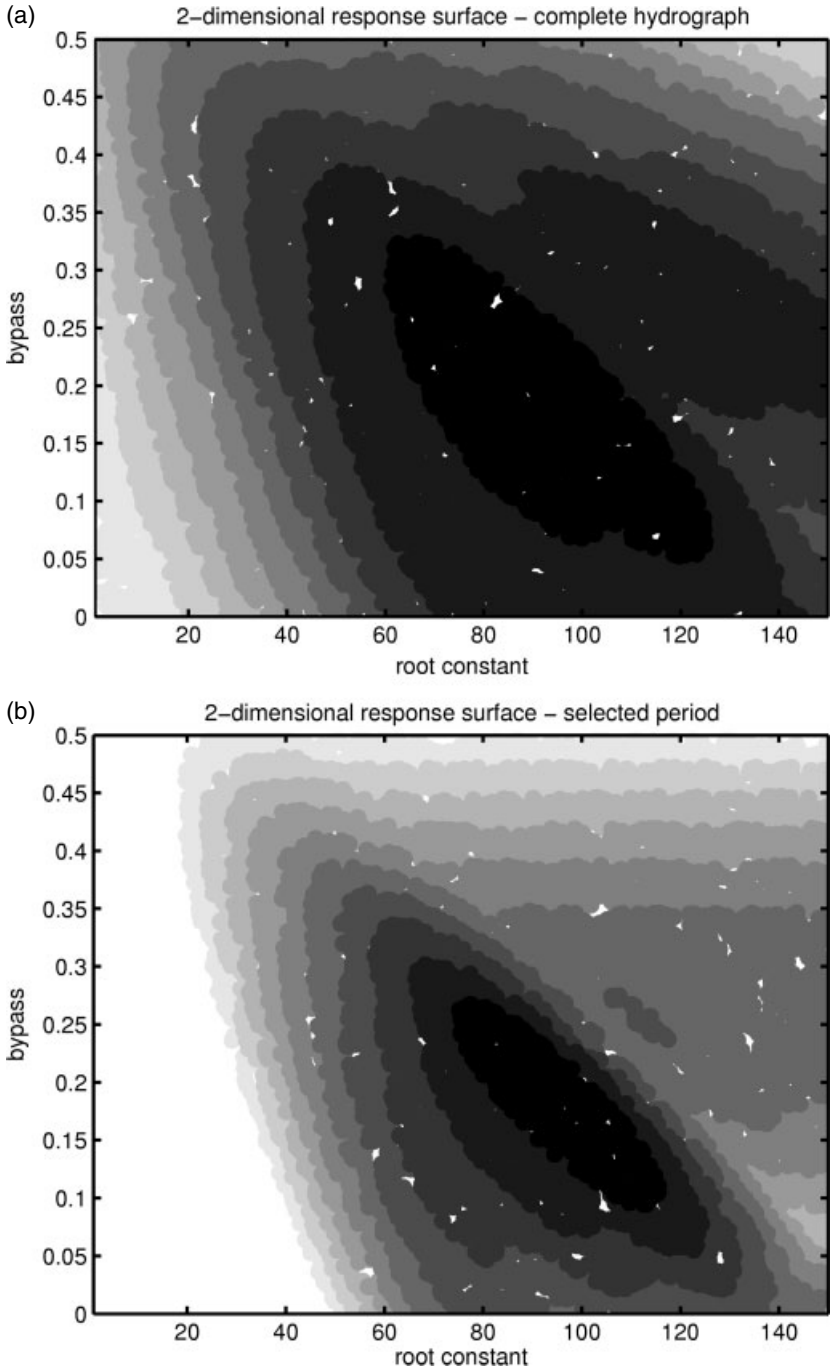

Figure 10. These two plots represent the response surface between the two parameters of the soil moisture accounting component, the root constant and the bypass. Both plots are based on a uniform random search sampling 10 000 points, during which the routing parameters were fixed to good performing values. Both plots consist of individual dots. The white areas are caused by a lack of density. The RMSE in plot (a) is calculated using the residuals from the complete time series, whereas plot (b) uses periods of high information content (see Figure 6), while avoiding the ambiguity of the bypass parameter identified in Figure 9. The time steps used in the selected period are 200 to 375 and 600 to 750. The smallest RMSE values are shown in black and the values increase by steps of 0·05 mm day^{-1} per contour

Inference and areas of possible model structure improvement

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

For example, the structural failure implied by the two distinct regions of preferred values of 'rc' (Figure 9) could be related to the fact that this component is required to fit the overall water balance and the wetting-up periods at the end of dry summers. A more flexible variant of this component, e.g. using a probability distribution of moisture stores in the catchment (Moore and Clarke, 1981; Moore, 1999), might perform better.

The analysis also indicates a failure with respect to the 'bypass' parameter. This parameter shows distinct areas of well-performing values in different parts of the feasible parameter space, as described earlier. Reasons for this could be that the process represented by the parameter is more complex than assumed, or that different processes, which could be represented separately, are aggregated into a single component. A possible improvement would be the replacement of this constant parameter with a dynamic component, e.g. by making the amount of precipitation contributing directly to the effective rainfall dependent on the soil moisture state of the catchment. This could account for features such as variable contributing areas, i.e. a larger part of the incoming rainfall contributes directly to the runoff when the catchment is very wet. However, further analysis is required to establish this.

CONCLUSIONS

The identifiability of parameters of dynamic and conceptual rainfall-runoff models is a difficult task. Manual calibration can yield good results, but the procedure is time consuming, requires experience, and does not allow for the objective analysis of parameter uncertainty and interaction. An automatic procedure overcoming these problems would, therefore, be highly advantageous. However, traditional, single-objective automatic calibration procedures often lead to a large number of similarly performing parameter sets, and to biased and therefore unacceptable model behaviour (Boyle *et al.*, 2000; Gupta, 2000).

Efficient and objective automatic procedures are required to allow for the evaluation and discrimination between competing models, i.e. parameter set and model structure combinations, and even between competing model structures. DYNIA is an attempt to develop an approach to complement traditional calibration methods to increase their discriminative power. The method is still in its testing stage, and more case studies will have to be performed.

Advantages of the method are its simplicity and its general utility (e.g. an application to a solute transport model can be found in Wagener *et al.* (2001b)). The computational effort involved, i.e. the Monte Carlo sampling procedure, is not a problem for most model structures with the computer power commonly available today. The number of simulations is, however, currently limited through the way DYNIA is implemented.

A list of possible areas of application of DYNIA is as follows:

1. To estimate parameters, i.e. simple model calibration or identification. The approach presented is an attempt to mimic manual calibration more closely than traditional schemes. One requirement of manual calibration is that the proper calibration of a conceptual model should result in parameters that cause model components to mimic processes they were designed to represent (NWS, 2000). This requires the isolation of the effects

of each parameter—a task difficult to achieve with manual or single-objective calibration, but in line with the DYNIA approach.
2. To analyse model structures. The awareness of the influence of model structural inadequacies on prediction uncertainty has grown in recent years, and the analysis of parameter variation in time is one possible approach to analyse this.
3. The algorithm relates model parameters and response modes of the natural system. The correct working of the model can therefore be checked. Do the parameters, and therefore model components, actually work as they are supposed to? Are there components that have little or no effect in producing the desired response?
4. To investigate data outliers or anomalies. Model structural error is not always the cause for time-varying parameter values. It can also be caused by erroneous data. Further analysis is usually required to be able to distinguish between the two. For example, data error might reveal itself by just being a one-off misfit between observed and calculated flow, whereas structural error will more probably be a consistent problem during similar response modes in different years.

There are also limitations and some possible further improvements to the methodology. The dependency on a Monte Carlo procedure can make it difficult to investigate high-dimensional parameter spaces. However, more complex model structures have yet to be analysed. A possible improvement could be to use the approach to find periods that contain information with respect to a specific parameter or parameter group in order to define objective functions to be used in a multi-objective optimization procedure. The first step of this application is shown in Figure 7.

It is important to realize that the philosophy behind this approach to structure identification and evaluation is not to validate or verify a selected model structure, an approach often considered not to be in line with a scientific approach (e.g. Oreskes *et al.*, 1994; Popper, 1999; Beven, 2001). The failure of a structural component should rather lead to the refutation (Popper, 1999) of the component and of the hypothesis underlying it. The modeller's task is then to develop an improved hypothesis based on the type of failure that has occurred using his knowledge and creativity (Beck, 1987). A model structure is (temporarily) accepted when no better-performing structure can be found and no underlying assumption is violated.

ACKNOWLEDGEMENTS

This project is funded by NERC under grant GR3/11653. We thank Southern Water for providing the data used in the example application. The constructive criticism of the two anonymous reviewers has led to improvements in the paper.

REFERENCES

Bard Y. 1974. *Non-linear Parameter Estimation*. Academic Press.

Bastidas LA. 1998. *Parameter estimation for hydrometeorological models using multi-criteria methods*. Unpublished PhD thesis, Department of Hydrology and Water Resources, The University of Arizona, USA.

Beck MB. 1985. Structures, failure, inference and prediction. In *Identification and System Parameter Estimation (Proceedings of 7th Symposium Volume 2, July 1985)*, Barker MA, Young PC (eds). IFAC/IFORS: York, UK; 1443–1448.

Beck MB. 1987. Water quality modelling: a review of the analysis of uncertainty. *Water Resources Research* **23**: 1393–1442.

Beven KJ. 1989. Changing ideas in hydrology—the case of physically-based models. *Journal of Hydrology* **105**: 157–172.

Beven KJ. 1998. *Generalized likelihood uncertainty estimation—user manual*. University of Lancaster, UK (unpublished).

Beven KJ. 2001. Towards an alternative blueprint for a physically-based digitally simulated hydrologic response modelling system. *Hydrological Processes* **6**: 189–206.

Beven KJ, Binley AM. 1992. The future of distributed models: model calibration and uncertainty in prediction. *Hydrological Processes* **6**: 279–298.

Boyle DP, Gupta HV, Sorooshian S. 2000. Towards improved calibration of hydrologic models: combining the strengths of manual and automatic methods. *Water Resources Research* **36**: 3663–3674.

Boyle DP, Gupta HV, Sorooshian S, Koren V, Zhang Z, Smith M. 2001. Towards improved streamflow forecasts: the value of semi-distributed modelling. *Water Resources Research* **37**: 2749–2759.

Dunne SM. 1999. Imposing constraints on parameter values of a conceptual hydrological model using baseflow response. *Hydrology and Earth System Sciences* **3**: 271–284.
Duan Q, Sorooshian S, Gupta VK. 1992. Effective and efficient global optimisation for conceptual rainfall-runoff models. *Water Resources Research* **28**: 1015–1031.
Ehrgott M. 2000. *Multicriteria Optimisation*. Springer-Verlag: Berlin.
Franks SW, Gineste P, Beven KJ, Merot P. 1998. On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas in the calibration process. *Water Resources Research* **34**: 787–797.
Freer J, Beven KJ, Ambroise B. 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resources Research* **32**: 2161–2173.
Goswami JC, Chan AK. 1999. *Fundamentals of Wavelets—Theory, Algorithms, and Applications*. John Wiley & Sons Inc.: New York.
Greenfield BJ. 1984. *The Thames catchment model*. Unpublished report, Environment Agency, UK.
Gupta HV. 2000. Penman Lecture. In *7th BHS National Symposium*, Newcastle-upon-Tyne, UK.
Gupta HV, Sorooshian S, Yapo PO. 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resources Research* **34**: 751–763.
Harlin J. 1991. Development of a process oriented calibration scheme for the HBV hydrological model. *Nordic Hydrology* **22**: 15–36.
Harvey JW, Wagner BJ. 2000. Quantifying hydrologic interactions between streams and their subsurface hyporheic zones. In *Streams and Ground Waters*, Jones JA, Mulholland PJ (eds). Academic Press: San Diego, USA; 3–44.
Hornberger GM, Spear RC. 1981. An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management* **12**: 7–18.
Hornberger GM, Beven KJ, Cosby BJ, Sappington DE. 1985. Shenandoah watershed study: calibration of the topography-based, variable contributing area hydrological model to a small forested catchment. *Water Resources Research* **21**: 1841–1850.
Jakeman AJ, Hornberger GM. 1993. How much complexity is warranted in a rainfall-runoff model? *Water Resources Research* **29**: 2637–2649.
Jolley TJ. 1995. *Large-scale hydrological modelling—the development and application of improved land-surface parameterisations for meteorological models*. Unpublished PhD thesis, Imperial College of Science, Technology and Medicine, London, UK.
Kleissen FM. 1990. *Uncertainty and identifiability in conceptual models of surface water acidification*. Unpublished PhD thesis, Imperial College of Science, Technology and Medicine, London, UK.
Kleissen FM, Beck MB, Wheater HS. 1990. The identifiability of conceptual hydrochemical models. *Water Resources Research* **26**: 2979–2992.
Kuczera G, Mroczkowski M. 1998. Assessment of hydrologic parameter uncertainty and the worth of data. *Water Resources Research* **34**: 1481–1489.
Legates DR, McCabe Jr GJ. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research* **35**: 233–241.
Moore RJ. 1999. Real-time flood forecasting systems: perspectives and prospects. In *Floods and Landslides: Integrated Risk Assessment*, Casale R, Margottini C (eds). Springer: Berlin; 147–189.
Moore RJ, Clarke RT. 1981. A distribution function approach to rainfall-runoff modelling. *Water Resources Research* **17**: 1367–1382.
Mous SLJ. 1993. Identification of the movement of water in unsaturated soils: the problem of identifiability of the model. *Journal of Hydrology* **143**: 153–167.
Mroczkowski M, Raper GP, Kuczera G. 1997. The quest for more powerful validation of conceptual catchment models. *Water Resources Research* **33**: 2325–2335.
Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, I, a discussion of principles. *Journal of Hydrology* **10**: 282–290.
NWS. 2001. *Calibration of the Sacramento model structure*. URL: http://hsp.nws.noaa.gov/oh/hrl/calb/workshop/parameter.htm. Accessed 15 February 2001.
Oreskes N, Shrader-Frechette K, Belite K. 1994. Verification, validation and confirmation of numerical models in the earth sciences. *Science* **263**: 641–646.
Penman HL. 1949. The dependence of transpiration on weather and soil conditions. *Journal of Soil Sciences* **1**: 74–89.
Piñol J, Beven KJ, Freer J. 1997. Modelling the hydrological response of Mediterranean catchments, Prades, Catalonia. The use of distributed models as aid to hypothesis formulation. *Hydrological Processes* **11**: 1287–1306.
Popper K. 1999. *The Logic of Scientific Discovery*. Routledge: London (First published 1959).
Sefton CEM, Howarth SM. 1998. Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. *Journal of Hydrology* **211**: 1–16.
Seibert J. 2000. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences* **4**: 215–224.
Seibert J, McDonnell J. 2001. Towards a better process representation of catchment hydrology in conceptual runoff modelling. Runoff generation and implications for river basin modelling. In *Freiburger Schriften zur Hydrologie*, Leibundgut C, Uhlenbrook S, McDonnell J (eds). Institute of Hydrology, University of Freiburg: Freiburg Germany. Band 13.
Sorooshian S, Gupta VK. 1985. The analysis of structural identifiability: theory and application to conceptual rainfall-runoff models. *Water Resources Research* **21**: 487–495.
Spear RC, Hornberger GM. 1980. Eutrophication in Peel Inlet, II, identification of critical uncertainties via generalised sensitivity analysis. *Water Research* **14**: 43–49.
Stigter JD, Beck MB, Gilbert RJ. 1997. Identification of model structure for photosynthesis and respiration of algal populations. *Water Science and Technology* **36**: 35–42.
Torrence C, Compo GP. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* **79**: 61–78.

Uhlenbrook S, Seibert J, Leibundgut C, Rohde A. 1999. Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. *Hydrological Sciences Journal* **44**: 779–797.
Van den Boogaard HFP, Ali MS, Mynett AE. 1998. Self organising feature maps for the analysis of hydrological and ecological data sets. In *Hydroinformatics'98*, Babovic V, Larsen LC (eds). Balkema: Rotterdam; 733–740.
Van Straten G, Keesman KJ. 1991. Uncertainty propagation and speculation in projective forecasts of environmental change: a lake-eutrophication example. *Journal of Forecasting* **10**: 163–190.
Wagener T. 1998. *Developing a knowledge-based system to support rainfall-runoff modelling*. MSc Thesis, Department of Civil Engineering, Delft University of Technology, Netherlands.
Wagener T, Lees MJ, Wheater HS. 1999. A generic rainfall-runoff modelling toolbox. *EOS. Transactions of the American Geophysical Union* **80**(17): F203.
Wagener T, Boyle DP, Lees MJ, Wheater HS, Gupta HV, Sorooshian S. 2001a. A framework for the development and application of hydrological models. *Hydrology Earth System Sciences* **5**: 13–26.
Wagener T, Lees MJ, Wheater HS. 2002. A framework for the development and application of parsimonious hydrological models. Mathematical Models of Large Watershed Hydrology, vol. 1, Singh VP, Frevert D (eds). *Water Resources Publications* LLC: USA; 91–140.
Wagener T, Camacho LA, Lees MJ, Wheater HS. 2001b. Dynamic parameter identifiability of a solute transport model. In Beheshti R (ed.). *Advances in Design Sciences and Technology—Proceedings of EuropIA'8*, Delft, April; 251–264.
Wagner BJ, Harvey JW. 1997. Experimental design for estimating parameters of rate-limited mass transfer: analysis of stream tracer studies. *Water Resources Research* **33**: 1731–1741.
Wheater HS, Bishop KH, Beck MB. 1986. The identification of conceptual hydrological models for surface water acidification. *Hydrological Processes* **1**: 89–109.
Wheater HS, Jakeman AJ, Beven KJ. 1993. Progress and directions in rainfall-runoff modeling. In *Modeling Change in Environmental Systems*, Jakeman AJ, Beck MB, McAleer MJ (eds). John Wiley & Sons: UK; 101–132.
Young PC, Beven KJ. 1994. Data-based mechanistic modelling and the rainfall-flow non-linearity. *Environmetrics* **5**: 335–363.
Young PC, Parkinson S, Lees MJ. 1996. Simplicity out of complexity in environmental modelling: Occham's razor revisited. *Journal of Applied Statistics* **23**: 165–210.

