Marcus Herbst

Dem Fachbereich VI
(Geographie/Geowissenschaften)
der Universität Trier
zur Verleihung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation

Using Self-Organizing Maps for the evaluation of watershed models

Betreuender:
Jun.-Prof. Dr.-Ing. Markus Casper

Berichterstattende:
Jun.-Prof. Dr.-Ing. Markus Casper
PD Dr. Thomas Udelhoven

Datum der wissenschaftlichen Aussprache:
16. Juli 2009

Trier, 2009

Acknowledgements

Contents

Appendix A

> Herbst, M., and Casper, M. C.: Towards model evaluation and identification using Self-Organizing Maps, Hydrol. Earth Syst. Sc., 12, 657-667, 2008.

Appendix B

> Herbst, M., Gupta, H. V., and Casper, M. C.: Mapping model behaviour using Self-Organizing Maps, Hydrol. Earth Syst. Sci., 13, 395-409, 2009.

Appendix C

> Herbst, M, Casper, M.C., Grundmann, J. and Buchholz, O.: Comparative analysis of model behaviour for flood prediction purposes using Self-Organizing Maps, Nat. Hazards Earth Syst. Sci., 9, 373-392, 2009.

Zusammenfassung

Für die Evaluierung als auch für die Optimierung hydrologischer Einzugsgebietsmodelle wird gewöhnlich auf statistische Abweichungsmaße zurückgegriffen, um den „Unterschied" zwischen der gemessenen und der simulierten Abflusszeitreihe zu quantifizieren. Die Reduktion des Informationsgehaltes von hydrologischen (Modell-) Zeitreihen durch die Verwendung dieser Abweichungsmaße ist jedoch, verglichen mit dem Umfang an Informationen, die man aus ihnen zum Zwecke der Modellidentifikation und –kalibrierung gewinnen möchte, sehr hoch. Es ist bekannt, dass der mit der Verwendung von aggregierenden Gütemaßen einhergehende Informationsverlust ein limitierender Faktor bei der Untersuchung von Modellen bzw. Modellzeitreihen ist und ein wichtiges Element der Modellunsicherheit darstellt, da sich sehr unterschiedliche Modellrealisationen mit nahezu identischem Gütemaß (z. B. r² oder RMSE) erzeugen lassen. Um den hierdurch induzierten Probleme bei der Modellidentifikation und Modelloptimierung zu begegnen, stellt die vorliegende Arbeit in drei aufeinander aufbauenden Studien einen Ansatz zur Modellevaluation vor, der auf der Verwendung von Selbstorganisierenden Merkmalskarten (SOM; Kohonen, 2001) basiert.

Bei SOM handelt es sich um ein künstliches neuronales Verfahren zur unüberwachten Klassifikation hochdimensionaler Daten, welches bisher sehr Erfolgreich zur Mustererkennung, wie z. B. bei der Sprachanalyse, eingesetzt wurde. Die zu untersuchenden Eingangsdatensätze werden dabei entsprechend der Ähnlichkeit in ihren Mustern auf benachbarten Positionen eines zweidimensionalen Rasters abgebildet. Die Neuronen genannten Elemente der SOM ergeben nach dem iterativen Training ein diskretes, d. h. datenkomprimiertes, Abbild der hochdimensionalen Eingangsdaten. Mittels verschiedener Visualisierungsmethoden lassen sich auf diese Weise Aufschlüsse über Häufigkeiten, Muster und Zusammenhänge in komplexen Datensätzen gewinnen. Ungeachtet ihres Potenzials haben SOM in der hydrologischen Modellierung, im Vergleich zu anderen Methoden künstlicher neuronaler Netze, nur wenig Beachtung erfahren. Die vorliegende Arbeit soll daher Einblicke und Anregungen bezüglich vorhandener als auch potenzieller Einsatzmöglichkeiten von SOM auf dem Gebiet der hydrologischen Modellierung liefern.

Die erste Studie (Appendix A) demonstriert, dass es möglich ist, die Ergebnisse einer Monte-Carlo Simulation mit einem Wasserhaushaltsmodell (4000 Simulationen mit jeweils über 17400 Zeitschritten) nach Ähnlichkeiten in ihren zeitlichen Mustern zu Ordnen. Hierbei kann sehr trennscharf zwischen verschiedenen Modellrealisationen unterschieden werden. Dies

versetzt den Anwender in die Lage, genaueren Aufschluss über bestimmte Aspekte des Verhaltens einzelner Parameter zu gewinnen, als dies z. B. mit Streudiagrammen möglich wäre.

Hierauf aufbauend, wird in der zweiten Studie (Appendix B) eine wesentliche Verbesserung des bestehenden Ansatzes eingeführt, die es ermöglicht, die Ergebnisse der SOM im hydrologischen Kontext zu interpretieren. Hierzu kommen die von Gupta et al. (2008) und Yilmaz et al. (2008) vorgeschlagenen Signature Indices zur Anwendung, welche den „Unterschied" zwischen simulierter und gemessener Zeitreihe in hydrologisch interpretierbarer Weise definieren. Dadurch ist es u. a. möglich, mittels SOM gestützter Auswertung Aussagen über die hydrologischen Funktionen der unterschiedlichen Modellparameter zu treffen, ohne Vorkenntnisse der Modellstruktur zu besitzen.

Den vorangehenden Ansatz aufgreifend, stellt die dritte Studie (Appendix C) eine vergleichende Analyse der hydrologischen Einzugsgebietsmodelle NASIM (Hydrotec, 2005), LARSIM (Bremicker, 2000) und WaSIM-ETH (Schulla und Jasper, 2001) vor. Unter Verwendung geeigneter Indices, welche Informationen über die Abbildung von Spitzenabflüssen extrahieren, wird untersucht, inwieweit die Modelle in der Lage sind, die Hochwasserereignisse in der gemessenen Zeitreihe zu reproduzieren. Hierbei wird insbesondere die Möglichkeit zu einer vergleichenden Auswertung der gewonnenen Daten ausgenutzt.

Die drei veröffentlichten Studien zeigen, dass man die Eigenschaften der SOM in Verbindung mit geeigneten Visualisierungsmethoden gezielt ausnutzen kann, um Informationen über bestimmte Aspekte des Verhaltens hydrologischer Modelle zu gewinnen. Neben der Gewinnung qualitativer Informationen über die Funktion und die Sensitivität von Modellparametern ist es vor allem möglich, diejenigen Modellergebnisse und Parametersätze zu identifizieren, welche die gemessene Abflussganglinie am Pegel, gemäß der jeweils angewendeten Kriterien, ‚am besten' repräsentieren. Dies kann, wie Appendix C zeigt, ggf. sogar modellübergreifend durchgeführt werden. Die im Rahmen der vorliegenden Arbeit gewonnenen Ergebnisse ermutigen dazu, weitere Untersuchungen zu möglichen Anwendungen von SOM in der hydrologischen Modellierung anzustellen.

Contributions of Self-Organizing Maps to an improved understanding of watershed models and dynamic systems

1 Introduction

1.1 The problem of watershed modelling

> *As far as the laws of mathematics refer to reality, they are not certain;*
> *and as far as they are certain, they do not refer to reality.*
> Albert Einstein

Models are simplified abstract representations which are devised to help in making statements about more or less tangible objects or processes. Model building as a progressive, iterative process founded on qualitative and/or quantitative observations which, depending on our understanding, starts with a perceptual model of the subject. Further formalization, which should also include the definition of system elements, boundaries and fluxes leads to the conceptual model that can take diverse shapes, e.g. verbal models or diagrams, however, this conceptual model still does not necessarily include the explicit relationship between the system elements. These come into play when the symbolic model is formulated by means of mathematical expressions which subsequently can be subjected to further reasoning and used in computerized numerical models. The latter finally allows for the making of testable deterministic predictions of a complex dynamical system like catchments (Gupta et al., 2008).

For the hydrological modelling of river catchments a broad range of deterministic simulation models are available, each of them offering different degrees of complexity in process description and spatial discretization. The different approaches to deterministic model building can generally be attributed to one of the following philosophies (Wagener et al., 2007; Schoups et al., 2008; cf. Klemeš, 1983):

- The *upward* or *reductionist approach*, which seeks to explicitly account for as much of the spatial structure and the known physical processes of the natural system as possible. Models that belong to this category "explain" the behaviour of the natural system and therefore are often referred to as "physically based" models.

- The *downward,* or *parsimonious* approach follows the principle of Occam's razor which advocates that a model should be not more complex than absolutely necessary to account for the behaviour of the integrated response of the natural system. Although, the governing equations of these *lumped conceptual* models have to obey the principles of mass- and energy conservation, this "effective" representation of the system behaviour does not provide an explanation of the processes. However, it is, in general, equally capable of satisfactorily predicting the system response such that the limitations of these models rather lie in their inability to use prior knowledge on the system (see e.g. Franchini and Pacciani, 1991; Gan and Biftu, 2003).

According to Brown and Heuvelink (2005, p.5) environmental models can be characterized as "inherently imperfect" which ultimately provides the reason for why the model structure is afflicted with uncertainty: Each deterministic environmental model, irrespective of its conceptualization and complexity, necessarily constitutes a simplified abstract representation of the natural dynamical processes. It has therefore to be stressed that the perceived "imperfectness" of environmental models is strongly relativized depending on the point of view that one assumes: In a use-oriented context, such as engineering, for example, a model is considered as a tool for solving certain problems. Here, any lack of accuracy in the model results is necessarily perceived as imperfectness. From a more scientific point of view, however, a model is nothing but the formulation of a hypothesis on the functioning and the description of a specific system. The confidence that can be attributed to its results acquire a different meaning and serve primarily as a means through which the underlying model hypothesis can be tested (Savanije, 2009).

The reductionist downward approach to environmental modelling is constrained by mathematical and computational tractability on one hand as well as the incompleteness of knowledge with respect to certain phenomena on the other hand (Klemeš, 1983). In practice, the mathematical tractability is alleviated with pertinent simplifications (see e.g. Dooge, 2005) while the gap of knowledge is often filled with "unverified assumptions" (Klemeš, 1983 p.7) which preferably refer to the homogeneity and time invariance of patterns and processes and ultimately undermine the physical plausibility of the model. Despite considerable advances, common process descriptions in hydrological modeling still do not account for many phenomena that may occur at the catchment scale, e.g. macropore flow or the release of mainly "old" water during storm events (Sivapalan, 2005). The difficulties of modeling hydrological processes at the catchment scale, however, obviously originate at a more fundamental level: Let alone the fact that many process theories and governing

2

equations that constitute our models essentially have been derived at the laboratory scale, it is important to recognize that catchment models are confronted with gaps between the scale at which the observations are made and the scale at which these processes actually occur (Sivapalan, 2005). In absence of a theoretical framework for lumping small scale processes for spatially heterogeneous units, it is merely an assumption that the physical equations as well as their parameters can be applied in identical manner across different scales (Beven, 1989). Despite all sophistication in model building, the bottom line is that in the presence of heterogeneity, self-organization, non-linearity and scale dependence of natural system processes current modeling approaches cannot establish the intended close physical reference to the natural system.[1] In consequence, "physically based models" can equally be treated as "lumped conceptual models" because their constituent equations and parameters inevitably have to remain "effective" representations of the system variables and processes (Beven, 1989, 2000). As any measurements to determine physically based parameters and boundary conditions tend to be devalued to point estimates that are likely to be invalid to represent a particular variable at the desired scale (Zehe and Blöschl, 2004; Kavetski et al. 2003) it is very hard – if not impossible – to define a unique best parameter *a priori*. A second reason for the non-uniqueness of model parameters also originates from the predominantly non-linear interactions between the coupled elements of the model structure as well as from uncertainties in the input data, including measurement error and noise. That way, at least when trying to solve the inverse problem for highly complex models, many degrees of freedom of the model are confronted with relatively few degrees of constraint, especially when the observation data is limited (Brown and Heuvelink, 2005).

In accordance with current standards, modelers attempt to solve this problem inversely by means of calibration, which is commonly based on the comparison of model generated input-state-output simulations with observed historical data (Sorooshian and Gupta, 1995). In this respect, the process of model calibration is linked to model evaluation, inasmuch as the analysis of the model output data is being used to draw conclusions about the model structure. In this regard, a model output which has a high degree of similarity to the observed system behaviour will always corroborate the model and its constituent hypothesis and vice-versa. The problem, however, lies in how such a comparison can be carried out and how its results can be interpreted (Gupta et al., 2008).

---

[1] In view of these fundamental difficulties, Sivapalan (2005) and similarly Troch et al. (2009), among others, argue that the current theory of catchment hydrology is facing an impasse. Hauhs and Lange (2008, 2006) come to the conclusion that our physical modeling paradigm is not capable of representing the complex and interactive behavior of natural systems, which is also attributed to biological interactions.

1.2 Model calibration and evaluation

In their simplest but most powerful form, calibration as well as the evaluation of a model is done visually because a trained expert is able to simultaneously discern various characteristics of the data and relate them to the hydrological context. This approach usually involves the comparative analysis of a number of different data characteristics that constitute the behaviour of the system and which, in turn, have to be satisfactorily reproduced by the model (i.e. a model and a corresponding set of parameters) in order to be acceptable ('behavioural'). (It is thereby often assumed that a model is suitable to represent the future behaviour of the system if its results are consistent with historical data. According to Oreskes and Belitz (2001), however, this widespread assertion is a fallacy because fitting the model to historical data might introduce a bias towards an existing trend. In particular, the assertion of 'validity' implies an unjustified legitimacy. Brown & Heuvelink (2005) point out that, ultimately, the model could still be 'right' for the wrong reasons). It follows that model calibration is inherently a multi-objective problem that requires at least the evaluation of more than one performance measure to ensure sufficient discriminatory power, even if the model produces only a single output time series (Gupta et al., 1998). Different strategies have been proposed to support and condition parameter identification by using complementary sources of information which can be classified according to Madsen (2003) as

- "multi-variable measurements", such as ground water measurements (Lamb et al., 1998) or combinations of different data sources (Ambroise et al., 1995; Gallart et al., 2007),

- "multi-site measurements", e.g. soil moisture observations (Franks et al., 1998),

- "multi-response modes" which evaluate the performance during different stages of the hydrological process using separate objective functions (Boyle et al., 2000).

Further, it is worth mentioning the work of Seibert and McDonnell (2003), who demonstrate how qualitative knowledge of the catchment behaviour can be included in the calibration process in order to yield a more realistic representation of the system. Interestingly, the inclusion of expert knowledge results in lower performance measure values (compared to a simple calibration against observed discharge) but better overall performance.

The multi-objective approach has already become a standard technique in optimization (Yapo et al., 1998; Madsen, 2003; Vrugt et al., 2003b) uncertainty assessment (Vrugt et al., 2003a) and evaluation (Gupta et al., 1998; Boyle et al., 2000; Wagener et al., 2003b) of hydrological

models. The use of multiple measures of performance implies that the optimum solution is not uniquely defined and consists of a Pareto set of parameters/model realizations that represent the different trade-offs among the objectives. Current multi-objective optimization techniques (Vrugt et al., 2003a, 2003b) effective and efficiently allow finding a Pareto optimal set of solutions. This provides valuable information on the capabilities of the model in reproducing the data. However, extracting the Pareto solution is not the only way that interesting insights into the model behaviour can be gained: By examining the model performance during different stages of the hydrological process (Boyle et al., 2000; Wagener et al., 2004) and particularly the temporal variability of parameter estimates (Wagener et al., 2003a; Beck, 1986) it became possible to relate model performance to individual model components. Since the inability of a model to reproduce the whole time series with a single parameter set may be indicative of structural model errors, this allowed testing the underlying model hypothesis (Gupta et al., 1998; Wagener et al., 2003a, Lin and Beck, 2007). The information regarding the functioning and the capabilities of the model finally provides guidance for potential model improvements.

An exciting new point of view for model evaluation also emerges from the application of Fourier and Wavelet transformation (e.g. Torrence and Compo, 1998) to model input and output time series. This approach particularly allowed to inspect the degree of model consistency in terms of temporal dynamics and scale dependence (e.g. Clemen, 1999; Labat et al., 2000; Lane, 2007; Montanari and Toth, 2007; Schaefli and Zehe, 2007).

Model evaluation is tightly related to the field of model uncertainty estimation. Current literature includes a plethora of techniques to address the different kinds of model uncertainty, ranging from pseudo-likelihood ratios (Beven and Binley, 1992), Bayesian methods (e.g. Thiemann, 2001; Vrugt et al., 2002; Kavetski et al., 2003; Misirli et al., 2003), Markov chain Monte-Carlo approaches (e.g. Kuczera and Parent, 1998; Vrugt et al., 2003a, 2003b), data assimilation techniques (Vrugt et al., 2005; Bulygina and Gupta, 2009), including particle filters and ensemble Kalman filters (Moradkhani et al., 2005; Weerts and El Serafy, 2006). Many of the aforementioned techniques are used in operational forecasting frameworks where they are geared towards minimizing the bias and the uncertainty in the model simulations (Gupta et al., 2008). Details concerning their application, however go beyond the scope of this work, which merely focuses on methods that help in the gaining of further insights into the patterns of model behaviour in order to test the underlying model hypothesis.

1.3 Inadequacy of statistical performance measures

Evaluation, as well as calibration methods that rely on a quantification of the 'difference' between the simulation and the measurements, commonly resort to regression based statistical measures of fit as the primary method of information extraction (Legates an McCabe Jr, 1999; Willmott et al., 1981; Nash and Sutcliffe, 1970). Hereafter, the terms 'performance measure' and 'objective function' will be used in identical manner to denote these measures. The pitfalls of using such measures are well known (e.g. see Hall, 2001; Lane, 2007; Schaefli and Gupta, 2007; Kavetski et al., 2003): Many objective functions imply assumptions about the statistical error properties that are often violated when they are used to quantify the degree of agreement between measured and simulated time series. This includes that the errors are either not normally distributed, do not have a mean of zero, show autocorrelation or heteroscedasticity. Kavetski et al. (2003) vividly demonstrates that the assumption of an inadequate error model can have devastating effects on parameter estimates. Most importantly, however, the nature of the 'information' we extract from the input-output data by means of these measures deserves critical attention: By putting data in the context of common statistical measures of fit, we allow primarily for a correlative evaluation of the data; e.g., the correlation coefficient informs about the percentage of the observed variance that can be explained by the model simulation, but conveys little or no information that relates directly to the hydrological context of the data (such as water balance, or velocity of the rainfall-runoff response). The use of statistics like the 'Nash Sutcliffe Efficiency' (Nash and Sutcliffe, 1970), for example, has virtually become a natural part of our daily modelling practice. However, we are often disregarding the fact that the 'Nash Sutcliffe Efficiency' measures the model performance relative to the observed mean value. The usefulness of the observed mean as a reference therefore varies strongly depending on whether the observed time series is characterized by fluctuations around a constant mean or by a strong seasonality. In any case it becomes evident that the 'Nash Sutcliffe Efficiency' can provide neither an absolute nor an interpretable measure of model performance. The majority of the quadratic error based performance measures[2] does not provide much better properties (Schaefli and Gupta, 2007).

To a certain extent, the choice of performance measures can be made such that the evaluation places a certain emphasis on particular parts of the hydrograph (Gupta et al., 1998). Even so, statistical measures of performance still have in common that the information which is

---

[2] An exhaustive list of currently used statistical performance measures in hydrological sciences is provided at the URL https://co-public.lboro.ac.uk/cocwd/HydroTest/index.html
The site offers the possibility for online comparison of measured and simulated time series.

potentially contained in a model time series as well as in its errors, is aggregated into a single numerical value, regardless of the characteristics of the time series and the actual pattern of the error. If, in order to sufficiently describe all relevant information which is contained in the model data, a vector of the dimension $\Re^{data} \gg 1$ is required, it follows that projecting the model data to a measure with the $\Re^{measure} = 1$ entails a substantial loss of information. The paradoxical attempt to determine a whole set of model parameters by using only one objective function or likelihood function clearly constitutes an ill-conditioned problem (Fig. 1) because, in addition to the abovementioned loss of information, a one-dimensional measure is insufficient to constrain a model with many degrees of freedom (Gupta et al., 2008).

Model output data dimension	Performance measure dimension	Parameter dimension
$\Re^{data}$	$\Re^{1}$	$\Re^{parameter}$
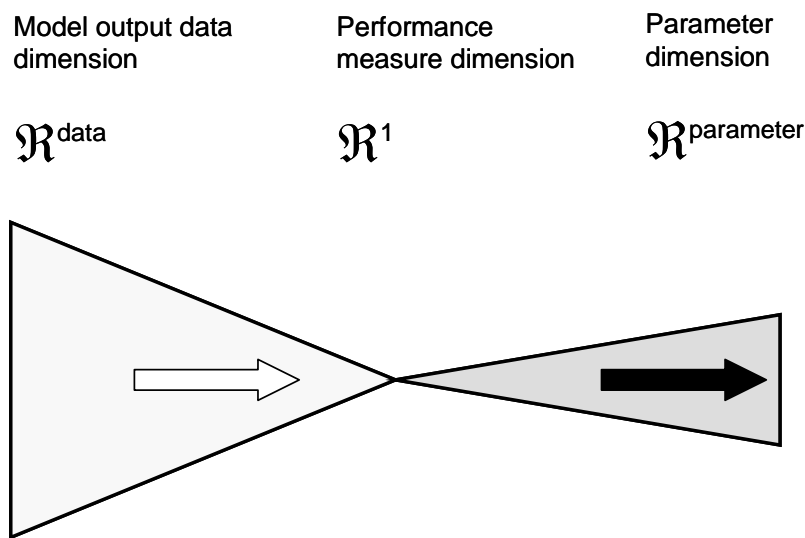
Fig. 1. Determining a set of model parameters using only one objective function (modified after Gupta et al. 2008).

In consequence, essentially different model results can be obtained with close to identical performance measure values such that model realizations may arguably appear to be equifinal[3], i.e. equally acceptable in reproducing the observed system behaviour (Beven and Binley, 1992; Beven and Freer, 2001; cf. Bertalanffy, 1969) as a consequence of the approach that has been used to discern them. The parameter sets that produce these equally acceptable results, however, will appear as non-unique and can be widely scattered throughout the parameter space (irrespective of the discriminatory power of the performance measure?).

---

[3] Long before the seminal paper by Beven and Binley (1992) the term „equifinality" was used by Ludwig von Bertalanffy in the context of General System Theory (Bertalanffy, 1969). He defines equifinality as "the tendency towards a characteristic final state from different initial states and in different ways based upon dynamic interaction in an open system attaining a steady state" (p. 46) as opposed to the behaviour of closed systems in which the final state is defined by the initial condition and the process (cf. pp. 131ff).

It is commonly accepted that this non-uniqueness of model parameters relates to model structural errors, input data uncertainty as well as measurement errors and –noise. However, it is as well important to recognize that the lack of discriminatory power of the measures which are used to compare simulated and observed time series potentially implies an additional drawback for the estimation of model parameters as well as for the analysis and evaluation of hydrological models (Lane, 2007; Wagener, et al., 2003b; Yapo, et al., 1998). The multi-objective approach to parameter estimation and optimization alleviates to some extent the problem which is related to the dimensionality of the data. However, commonly used statistics do not provide the information which is needed for a diagnostic evaluation of model behaviour (Gupta et al., 2008).

Diagnostic model evaluation is a fundamental step in the iterative model building and identification process that requires adequate tools that guide the modeller towards targeting the causes of 'unsatisfactory' model behaviour. Because variance-based statistical measures fail to extract and relate the information contained in the model time series data to characteristics that are interpretable and meaningful in the context of the hydrological theory, they offer only minimal diagnostic support when used at the front end of various model evaluation frameworks. In response to this, Gupta et al. (2008) and Yilmaz et al. (2008) recently proposed the concept of a diagnostic evaluation approach rooted in information theory, in which a set of measures is used to characterize various theoretically relevant system functions and process behaviours. The term "Signature Indices" is introduced to distinguish these measures from conventional variance based performance measures that lack relationships to the underlying hydrological theory. Yilmaz et al. (2008) propose a set of such Signature Indices to help guide the diagnostic process of model evaluation in a meaningful and interpretable way, inasmuch as these measures correspond to four major hydrological functions of the watershed: overall water balance, vertical redistribution, temporal redistribution and spatial redistribution.

1.4 Confronting high-dimensional data distributions

Because the evaluation of environmental and hydrological models is inevitably confronted with non-uniqueness, model realizations (i.e. parameters) and their corresponding objective function values have to be represented using probability distributions. However, in view of the often highly non-linear interdependences between model parameters that include compensating effects and threshold behaviour, the possibilities of classical statistical methods of system identification are often limited. In addition, these methods are subject to many

8

statistical assumptions that mostly do not apply (Wagener et al., 2001). Alternatively, Monte-Carlo simulation provides a robust method to gather information on the model response with respect to variations of the parameter (Melching, 1995). This method uses repetitive model simulation runs on the basis of randomly sampled parameters in order to determine an empirical probability distribution of the model output. In cases where no prior knowledge on the distribution of feasible parameters is available, a uniform random distribution within plausible ranges has to be assumed. The problem with Monte-Carlo simulations is the number of model runs which theoretically would be required to attain a statistically satisfying sampling density of the parameter space. The reason is that the region of interest in the parameter space with dimension n can be seen as a hypercube with the side length l and the volume l^n. Substituting the side length with the number of samples we deem appropriate along each axis of the parameter space and n with the number of model parameters it becomes evident that, in the vast majority of studies on model performance, the required number of model runs becomes prohibitive, even if assuming a run-time of only one second per model execution. More advanced approaches (e.g. Kuczera and Parent, 1998) exist that resort to Markov Chain Monte Carlo (MCMC) methods in order to approximate an unknown distribution (e.g. parameter distributions) by means of stochastic simulation. The most widely used building block herein is the Metropolis-Hastings algorithm (Metropolis, 1953; Hastings, 1970) which requires prior definition of a proposal distribution for the sampling. With a poor choice of this initial distribution, however, the algorithm looses its efficiency and converges only slowly which often makes its application to environmental models, where a priori knowledge on distributions is generally scarce, somewhat difficult (Vrugt et al., 2003b). Vrugt et al. (2003b) developed an optimization algorithm that overcomes these problems by using an adaptive MCMC scheme.

A common method to analyze a Monte-Carlo simulation with a hydrological model is to calculate a set of objective functions for each model output time series relative to some historical observation (here it is assumed that only one time series per model execution is examined). As mentioned above, this is equivalent to projecting the time series to a lower dimensional 'objective function space' (assuming that the number of objective functions is much smaller than the number of elements in each time series vector). In a one- or two-dimensional case the results can easily be visualized using scatterplots and response surfaces, respectively. However, if three or more objective functions are used one commonly has to resort to orthogonal projections to individual axes or planes of the objective function space and the results get much more difficult to comprehend and to visualize. Due to the non-linear

behaviour of dynamical models of natural systems and the limitations inherent to the observed data, the resulting response surface generally eludes any mathematical description. For this reason, considerable research has been directed towards algorithms that allow finding the global optimum of response surfaces from complex environmental models (for an overview see Sorooshian and Gupta, 1995; Gupta et al., 2003; Duan et al., 2003). The difficulties in developing such automated calibration techniques originate from the following properties of the model response surface (Duan et al., 1993; Sorooshian and Gupta, 1995):

- There may be more than one main region of convergence.

- Each of these regions may contain possibly uncountable local optima at various distances from the best solution. Optima can be found at multiple scales.

- The model response surface may not be smooth nor continuous.

- The parameters may show highly nonlinear interdependence as well as varying degrees of sensitivity.

- The model response surface near the optimum is often non-convex.

Considering these characteristics it becomes clear that solving this optimization problem may be difficult or even impossible using local or derivative based search methods and that it may be impossible to obtain a unique solution of optimal parameters. With the Shuffled Complex Evolution algorithm (SCE-UA) Duan et al. (1992) presented an effective and efficient global optimization method. Its success in a wide range of applications can be attributed to a hybrid strategy that combines elements of deterministic, evolutionary and stochastic exploration of the model response surface which ultimately demonstrates that the key to an effective automatic calibration technique lies in finding an effective way to *gather* and *use information* from the model response surface.

Similarly, model evaluation is confronted with the multi-dimensionality and complexity of the model response. In the first place, as mentioned above, model evaluation as well as automatic calibration rely on a projection of the original model output time series data to a lower dimensional 'objective function space'. This first, basic step defines the context in which the model outputs are analyzed while at the same time performing a lossy compression of the data which is intended to exclusively preserve what is assumed to be the important information in it (Gupta et al., 2008). Although the focus of model evaluation is a different one, the example of the Shuffled Complex Evolution algorithm also illustrates what lies at the heart of the model identification and model evaluation problem: As the original model response – and the

projection of the model output into an objective function space respectively – is multidimensional it eludes any analytical description that could be used for further reasoning on model performances and trade-offs between different modelling objectives. Hence, model evaluation is essentially concerned with *gathering* and *organizing* the information contained in the entire multidimensional data distribution that constitutes the output from a high number of model runs. In order to make the results comprehensible to the human being, a further, essential component of model evaluation is the *visualization* of the extracted information.

1.5   Underlying hypothesis

As exposed in the preceding sections, the properties of environmental models and the data produced by them suggest that a major key towards improving our abilities to gain insight into the patterns of model behaviour with the aim of testing the underlying model hypothesis may lie in

- enhancing the abilities to discriminate between different model output time series (as commonly used aggregating statistics lead to ambiguous results),

- using meaningful similarity measures that help in making inferences about the model structure (as statistical measures do not refer to the hydrological context),

- Finding effective ways of capturing and visualizing complex high-dimensional data distributions (as they potentially may provide all relevant information on the model behaviour).

1.6   Goals of this study

Based on three consecutive studies the present work proposes an alternative approach towards hydrological model evaluation that aims at taking into account the goals that have been formulated in Sect. 1.5. A central position in these studies is occupied by the application of Self-Organizing Maps (SOM; Kohonen, 2001) which is used to represent and to analyze high-dimensional data obtained from Monte-Carlo simulations with distributed hydrological watershed models.

The aim of the present work is to demonstrate that the application of Self-Organizing Maps has very high potential to address fundamental issues of model evaluation: It is shown that the clustering and classification of model time series by means of SOM can provide useful insights into model behaviour. In combination with the diagnostic properties of Signature Indices (Gupta et al., 2008; Yilmaz et al., 2008) SOM provides a novel tool for interpreting

the model parameters in the hydrological context and identifying parameter sets that simultaneously meet multiple objectives, even if the corresponding model realizations belong to different models. In this context, it is interesting to note that the SOM based method presented in this work in some aspects aims at similar objectives as the Parameter Identification Method based on the Localization of Information (PIMLI) developed by Vrugt et al. (2001; 2002). The PIMLI method attempts to find disjunctive subsets of the observations that contain the most information for the different model parameters. PIMLI thus provides a diagnostic instrument that helps in differentiating between parameters and model concepts. In contrast to the SOM based method, PIMLI approaches the model evaluation problem more from the observations than the model itself. It uses a sequential optimization methodology that merges the ideas of generalized sensitivity analysis (GSA; Spear and Hornberger, 1980), Bayesian recursive estimation (BARE; Thiemann et al., 2001) and the Metropolis algorithm (Metropolis et al., 1953) for efficient sampling.

2   Self-Organizing Maps and their application in hydrological sciences

A Self-Organizing Map (SOM; Kohonen, 2001) consists of an unsupervised learning neural network algorithm which is applied to extract an "archetypal" (Sang et al., 2008, p. 1194) set of characteristics that describe a high-dimensional data set. In the process, it performs a nonlinear projection of high-dimensional input-data onto a lower dimensional plane. Thereby, in a sense, it mimics a working principle of the cerebral cortex (and other structures of the brain) where multi-dimensional stimuli are assigned to two-dimensional arrays of neurons in a topology preserving manner, i.e. similar stimuli are mapped onto nearby regions. The stimuli that are presented to the SOM, however, consist of data vectors from a larger data set, which is referred to in the following as the training data set. SOM is capable of discerning and categorizing the range of data patterns that occur in the training data which render it a popular tool in exploratory data analysis (Kaski, 1997; Vesanto, 2000). Unlike the more common artificial neural network techniques, as for example feed-forward multilayer perceptrons (MLP, for a comprehensive overview of current artificial neural network approaches the reader is referred to Haykin, 1999) SOM has no output function and is therefore conceived as a clustering and data-compression technique. In an iterative training procedure SOM produces a smooth mapping of vectorial input data items on a discrete grid of map units which are also referred to as neurons or nodes. Commonly the neurons are arranged on a rectangular, two-dimensional grid with hexagonal topology (Fig. 2). This hexagonal configuration has the

advantage that, in contrast to a rectangular topology, the lateral distances towards each neighbouring neuron are equal.

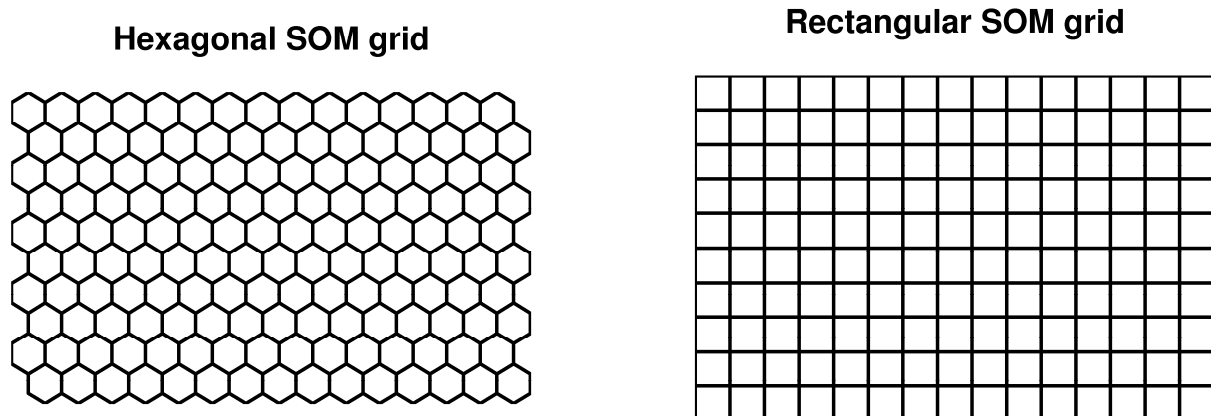**Hexagonal SOM grid**   **Rectangular SOM grid**

Fig. 2. Configurations of the SOM grid.

Depending on the scope of the SOM application, it is also possible to use other shapes than a rectangular, e.g. a toroidal grid. Analogous to the functioning of the cerebral cortex, nearby locations on this mapping are attributed similar data patterns. Most notably, in contrast to clustering methods, such as k-means clustering (see Kohonen, 2001), SOM is topology preserving – a distinctive feature of the cerebral cortex – which is why similar patterns are projected onto contiguous locations. Accordingly, the distance between two nodes on the mapping is proportional to the dissimilarity of the data items attributed to them. However, because the number of neurons on the map is always inferior to the number of data items that are presented to the map in the course of the training, each neuron is attributed a small subset of the training data set with similar properties. Note that each neuron is characterized by a reference- or codebook vector that can be interpreted as the cluster centroid of the data items that were attributed to its corresponding map unit during the training. At this point, SOM is closely related to vector quantization (see Kohonen, 2001) and data-compression techniques as through its reference vectors a SOM provides a discrete approximation of the high-dimensional distribution of the training data. The application of an SOM usually comprises four consecutive steps:

- *Normalization of the input data:* Prior to the training the input data has to be normalized in order to avoid that variables with predominantly high values exert a disproportionate influence compared to other variables. Most commonly the variables are scaled to values having zero mean and variance of one. (see Appendix B, Sect. 2.4).

- *Dimensioning and initialization:* Generally dimensions of the map, i.e. the side lengths and the number of neurons on a map, can be freely adjusted to the requirements of the particular application. In many cases, however, an heuristic procedure (Vesanto et al., 2000) is used where the number of neurons on the map is $m = 5\sqrt{N}$, with N being the number of items in the training data. Further, the side lengths of the map are determined based on the ratio of the two biggest eigenvalues of the training data covariance matrix (see Appendix B, Sect. 2.4).

- *Iterative training:* During the iterative training of an SOM each vectorial input data item, which is picked randomly from the training data set, is attributed to the neuron whose reference vector has the smallest Euclidean distance from it. This neuron is referred to as the "best-matching-unit" (BMU). Subsequently, depending on their distance from the BMU and the number of elapsed training cycles, the reference vectors in the neighbourhood are updated. These consecutive steps of each training cycle represent a competitive and an adaptive learning process. The mapping finally "self-organizes" and converges upon repeated cycling through the input data set. The training process is explained in detail in Sect. 2.1 of Appendix A and in Sect. 2.3 of Appendix B.

- *Post-processing:* A wide range of techniques is available that help with the comprehending of the resulting map by visualizing its reference vectors in different ways. This way, the clustering structure, the data distribution or correlations between individual variables can be illustrated. A concise review of the existing methods is provided by Vesanto (1999). However, depending on the scope of the SOM application, the post-processing of the map can be extended to subsequent cluster analysis of reference vector properties (e.g. Obayashi and Sasaki, 2003; Reusser et al., 2008), derivation of probability density functions (Gutiérrez et al., 2005; Kim et al., 2006), likelihood estimation (Kim et al., 2006) or local modelling (Vesanto, 1997; Principe et al., 1998; Hsu et al., 2002). An SOM may also be used in hybrid approaches along with other ANN where it serves as a pre-processor for clustering the input data domain (Abrahart and See, 2000; Moradkhani, 2004; Jain and Srinivasulu, 2006)

The most notable properties of the mappings produced by the SOM algorithm can be summarized as follows:

- The SOM provides a discrete approximation of the input data space such that the reference vectors follow the distribution of the input data. As the reference vectors represent local conditional expected values of the data the SOM can be seen as essentially equivalent to a discretized form of principal curves (Kaski, 1997; Haykin, 1999), i.e. a generalization of the principal component analysis. Principal component analysis (PCA) can be used to determine the axes of a linear sub-space that preserve a maximum of the variance of the data. However, as these axes are linear, PCA cannot account properly for curved or arbitrarily shaped data distributions which are frequently encountered e.g. when evaluating the results of Monte-Carlo experiments with environmental models. In conjunction with SOM PCA is often used to visualize the training process and to show how the reference vectors adapt to the distribution of the training data (Fig. 3). Principle curves constitute a generalized form of PCA: Instead of projecting the data into a linear subspace, the principle curves span a non-linear sub-space using smooth curves that are formed by the mean value of all data points that project to it. Thus, each point on a principal curve constitutes a conditional expectation value.

- The topology preservation has the advantage that it allows to see the data which is used for the training in context.

- Patterns in the input data that occur more frequently are mapped onto a larger area of the map.

- The SOM has visually appealing properties that facilitates the comprehension of the data.

Oja et al. (2002) reports that from the introduction of the SOM algorithm in 1981 until 2001 more than 5300 papers have been published that either apply the algorithm, have benefited from SOM or contain an analysis of it. Within the period of 1988 until 1999 Oja et al. (2002) predominantly recorded diverse applications in fields such as image and video processing, pattern recognition, mathematical applications, artificial intelligence, engineering in biology and medicine, coding, speech, control, signal processing, circuits, documentation as well as business and administration. Since 2001 more than 2000 additional papers on SOM have been published[4], this may also include findings from other disciplines.

---

[4] The bibliography is available in BibTeX format at the URL http://www.cis.hut.fi/nnrc/refs/
Following this reference, 7718 works have been based on SOM until 30.05.2007 (last update of the page).

Although applications of artificial neural networks (ANN) are already subject to intensive research in environmental modelling and the water resources field (see Minns and Hall (2005) as well as Govindaraju and Ramachandra Rao (2000) for an overview) it can well be assumed that SOM has not yet found this widespread popularity. Yet, numerous possibilities for applications are evident which is also exemplified by the existing applications of SOM in hydrology in neighbouring disciplines. In this regard, Kalteh et al. (2008) report a steady increase in the use of SOM in the water resources area in the recent years which, last but not least, is also due to the robustness and the outlier- as well as noise tolerant properties of the method (Allinson and Yin, 1999).
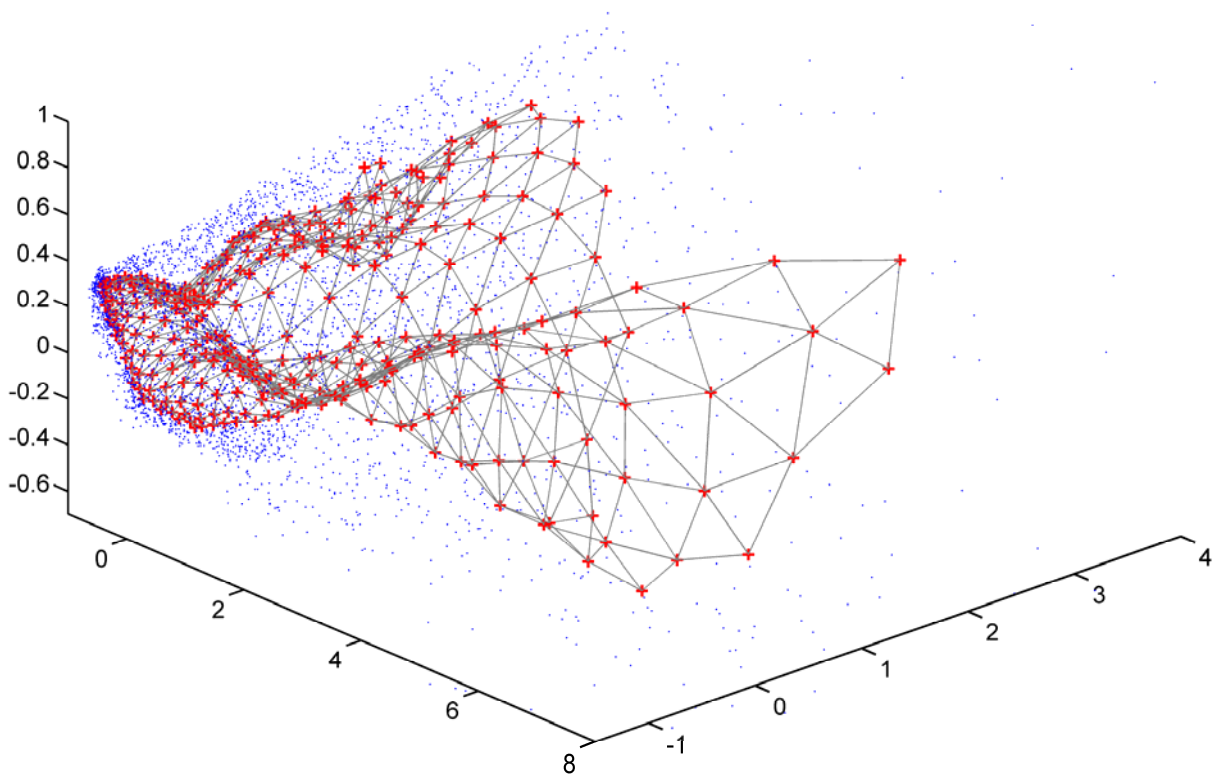
Fig. 3. Visualization of reference vector positions (+) in relation to the training data.

In a rather general sense, the SOM is one of several methods that can be used for time series clustering (Warren Liao, 2005), process monitoring (Alhoniemi et al., 1999; Simula et al., 1999), time series forecasting (Simon et al., 2005) or local time series modelling, as already mentioned above.

In one of the earliest applications to hydrological issues Boogaard et al. (1998) explored the clustering of discharge patterns of the Mekong river based on monthly averaged discharges from 1925 to 1986 on a 6×6 SOM grid. Their preliminary study unexpectedly resulted in a

physically meaningful classification of different annual discharge characteristics. They further applied the SOM successfully to analyze daily records of ecological and hydrochemical parameters which allowed for an ecological classification of 9 lakes. Similarly, Lischeid (2003, 2006) reports a very successful application of an SOM to analyze spatial and temporal trends in hydrochemical groundwater data. As a consequence, the approach is being used within the groundwater quality monitoring network of the Bavarian Geological Survey. In a similar context, Peeters et al. (2007) developed a new variant of the SOM algorithm (GEO3DSOM) which allowed to explicitly incorporating three-dimensional knowledge on the data.

SOM-based analysis concisely captures the spatial and temporal variability in climatological phenomena and lends itself for a wide range of applications in meteorology and synoptic climatology (see also Hewitson and Crane, 2002; Sang et al., 2008). Lin and Chen (2006) apply an SOM to precipitation data from 154 rain gauges in Taiwan, with a total of 5920 annual rainfall observations, in order to identify homogeneous regions for regional frequency analysis. Based on 17 input variables (gauge latitude, gauge longitude, elevation, mean annual rainfall, standard deviation of annual rainfall and mean monthly rainfall for every month) and a map with 12×12 neurons they obtained eight homogeneous regions. They further compared the SOM results with the k-means and Ward's clustering method and conclude that these techniques were outperformed by the SOM approach in terms of membership accuracy. Kalteh and Berndtsson (2007) explore the use of an SOM both for the regionalization and interpolation of precipitation in northern Iran, a region with considerable complexity in the precipitation mechanism. In their approach, classification of rainfall data is performed using an unsupervised SOM whereas for the subsequent interpolation a supervised SOM is applied. They conclude that the SOM is capable of finding regions with similar precipitation mechanisms and can be used to accurately interpolate rainfall data. Gutierrez et al. (2005) present an interesting application from the field of meteorology: They use SOM to characterize atmospheric patterns for the downscaling of atmosphere-ocean multi-model ensemble forecasts. To this end, probability density functions (PDF) for particular phenomena are derived directly from the SOM of climatological reanalysis data which are subsequently combined with a representation of local precipitation regime. Thus a probabilistic numerical precipitation forecast for a given station is achieved using PDF derived from a SOM of climatological reanalysis data. Using these PDF they were further able to derive an entropy-based measure of the predictability for the downscaled forecast.

In the context of watershed- as well as land surface modelling various researchers achieved remarkable results using the SOM in hybrid approaches where it is being used to divide the input data domain into distinct regions of event types which are subsequently represented by separate models. This way, Abrahart and See (2000) performed a grouping of hydrologic event types based on six hours of previous flow data and an SOM consisting of 8 × 8 map units which was used to implement a multi-model approach using separate ANN for different clusters. The results outperformed an ANN which was run on the entire data and an autoregressive moving average (ARMA) model. Similarly, Jain and Srinivasulu (2006) explored the use of a one-dimensional SOM model for decomposing the rainfall-runoff data space into different classes. The SOM is subsequently coupled to different ANN (feed-forward MLPs) as well as conceptual approaches in order to model the individual, fragmented rainfall-runoff relationships. One of the most interesting applications in this area is the Self-Organizing Linear Output mapping network (SOLO) developed by Hsu et al. (2002). The SOLO uses a SOM-based classification of rainfall-runoff patterns in combination with local linear regressions. The six input variables for the training of a 15 × 15 SOM consisted of area averaged rainfall data for three days along with the corresponding discharge measurements. After the training was accomplished the authors found that the resulting map provided insight into the different stages of the rainfall-runoff process in the catchment such that different regions on the map represented meaningful temporal characteristics of the process which could be classified into base flow region, increasing rainfall region, peaking hydrograph region, quick recession and slow recession region (cf. Boyle et al., 2000). They fitted separate piece-wise linear regression functions for the data projected onto the individual neurons which was used to perform one-day ahead predictions of streamflow. The performance of SOLO was subsequently compared to an autoregressive model with exogenous inputs (ARX), a multi-layer feed-forward network (MFN), a recurrent neural network (RNN) and to a conceptual rainfall-runoff model (SAC-SMA). The authors found that the SOLO as well as the MFN were capable of outperforming the remaining approaches. More recently, Abramowitz et al. (2006, 2007) used the SOLO approach by Hsu et al. (2002) to characterize and quantify the systematic output error of a land surface model as a function of its meteorological input data. Similarly, Abramowitz et al. (2008) applied an SOM to cluster time steps with similar meteorological condition in order to examine the conditional bias of land surface models; and to evaluate multi-model independence (Abramowitz and Gupta, 2008). Another hybrid architecture was presented by Moradkhani et al. (2004) who developed the Self-Organizing Radial Basis (SORB) function for one-step ahead forecasting of daily

18

streamflow. SORB primarily resorts to a Gaussian Radial Basis function (RBF, see Haykin, 1999; Govindaraju and Zhang, 2000). The center and spread parameters, however, which are required for the Gaussian transfer function were derived using an SOM. This is because, according to the principles of SOM, the reference vectors represent local conditional expectations and can thus be regarded as centers. Further, the spread parameters were calculated from the mapping based on the density of data items that were attributed to each map unit, i.e. the distance of all data items from its corresponding reference vector. (Note that, this way, Moradkhani et al. (2004) successfully capitalized on the statistical properties of the reference vectors). In terms of forecasting accuracy, SORB was found to provide superior performance compared to a linear regression model, a feed-forward MLP and a SOLO (Hsu et al., 2002).

Apart from rainfall-runoff modelling itself, SOM has also been successfully applied in the context of modelling or inferring physical and hydrological properties of catchments. Schütze et al. (2005) developed an enhanced Self-Organizing map architecture that includes multiple input-output relationships (entitled SOM-MIO) and, based on a linear interpolation scheme, the possibility to generate continuous outputs. The SOM-MIO proved capable of approximating the Richards equation and even performed its inverse solution with outstanding accuracy. The authors report that the different tasks did not require additional training and that, in simulating a laboratory irrigation experiment, the SOM-MIO demonstrated improved performance compared to a numerical simulation model. Further, Huang et al. (2003) present a framework for transferring model parameters from data-rich areas to data sparse areas which is illustrated with the b-parameter of the Variable Infiltration Capacity (VIC) model. Their methodology is based on a two-stage approach: In a first step a SOM and successive k-means clustering is used to categorize soil data sets. Subsequently a supervised neural network along with a Bayesian regularization scheme is applied to each of these clusters in order to establish nonlinear relationships between the soil information and model parameters. Islam and Kothari (2000) introduce a methodology based on SOM to characterize the spatial distribution of soil moisture and soil texture from remote sensing data. The spatial structure of 21,204 soil moisture measurements with a spatial resolution of 200 m is represented in a map of 49 neurons. The authors highlight that this mapping is capable of preserving the second order statistical properties of the data and that the preservation of spatial information in the map allows reconstructing the soil moisture map at different scales of aggregation. They further exploit the relationship between the space-time evolution of brightness temperature data and soil types in order to identify soil types using SOM and a sequence of remotely sensed

images. In this approach the input data for the SOM consists of (irregular) time series of brightness temperature data for each pixel. Although, in their study, the number of neurons used for the SOM is based on the number of known soil classes, Islam and Kothari (2000) point out that their approach does not necessarily rely on this a priori knowledge. The trained SOM of only three nodes clusters the brightness temperature time series into distinct categories that represent the archetypal dry-down characteristics of different soil textural classes. A comparison of their SOM generated map of soil textural classes revealed very close correspondence to the observations. A similar approach was later used by Chang (2001) to infer soil hydraulic properties. With the aim of developing practical measures for soil quality Mele and Crowley (2008) use SOM to examine the interrelationship between various biochemical signatures of microbial communities, genetic signatures and various soil chemical, physical and biological parameters. They conclude that SOM provides an intuitive means that could assist land managers to examine and interpret complex relationships in soil analysis data.

The abovementioned studies indicate that SOM can serve as a valuable tool in a variety of stand-alone and hybrid approaches for modelling the patterns or dynamics of environmental systems. Far less attention, however, has been directed towards applying SOM for the evaluation of the complex behaviour of deterministic simulation models. Reusser et al. (2008) demonstrate an interesting approach in this field which, apart from Appendix A, B and C, at present remains the only publication in this field. They analyze the temporal dynamics as well as the type of model error by using an SOM for the assessment of multiple performance measures from a moving window approach (cf. Wagener et al., 2003a). The assessment resorts to a set of performance measures which are calculated for a time-variable window, yielding a vector of performance measures for each time step. Subsequent thinning of these vectors is carried out in order to remove measures that show temporal correlation. The remaining measures are interpreted using synthetic peak error functions and SOM for further data reduction and classification of error types. Although impressive diagnostic possibilities of this approach have been demonstrated, the interpretation of the results appears to be rather cumbersome.

3   Three studies towards hydrological model evaluation using Self-Organizing Maps

In the three studies presented in this work a SOM is used to obtain a topologically ordered classification and clustering of model generated time series.

20

In the first study (Appendix A[5]) SOM is used to differentiate the spectrum of simulated time series obtained from Monte-Carlo simulations with the distributed conceptual watershed model NASIM (Hydrotec, 2005). This first contribution is primarily geared towards circumventing the lack of discriminatory power which is inherent to aggregating statistical performance measures. What is therefore explored is whether a SOM can be applied to yield a topologically ordered mapping of model output time series according to the similarity in the temporal patterns of their residuals. To this end, the properties of the SOM mapping are examined by relating its elements to statistical performance measures and to the parameter values that have been used to generate the model results. Further, the SOM is used to identify those model realizations among the set of Monte-Carlo realizations that most closely approximate the temporal pattern of the measured time series. The result is analyzed and compared with the manually calibrated model as well as with the result of a single-objective Shuffled Complex Evolution algorithm (SCE-UA, Duan et al., 1992) which is used to minimize the RMSE. The results confirm that it is possible to successfully train an SOM even if the dimensionality of the individual data items is very high. In the present case, each input data item (and consequently each reference vector) represents a time series consisting of more than 17,400 time steps. It is illustrated that, irrespective of this high dimensionality, very individual properties of the training data time series were attributed to each element of the resulting 22×15 SOM such that the mapping preserves the correlation structure of various statistical measures of performance. It is further shown that, the proposed SOM application is capable of revealing information about parameter sensitivities as well as parameter interactions that has not been available when using scatter plots for model evaluation. The model realizations obtained by using the same mapping in order to identify the model realizations that most closely approximate the observations largely outperformed the result of the SCE-UA application as well as the manual calibration. These findings provide another indication of the high discriminatory power which is supported by the Self-Organizing map and thus encourages its further use for model evaluation purposes.

Based on the same data, the second study Appendix B[6] proposes an improvement to the previous approach. This is achieved by linking the SOM to the abovementioned Signature

---

[5] Herbst, M., and Casper, M. C.: Towards model evaluation and identification using Self-Organizing Maps, Hydrology and Earth System Sciences, 12, 657-667, 2008.
http://www.hydrol-earth-syst-sci.net/12/657/2008/hess-12-657-2008.pdf

[6] Herbst, M., Gupta, H. V., and Casper, M. C.: Mapping model behaviour using Self Organizing Maps, Hydrology and Earth System Sciences, 13, 395-409, 2009.
http://www.hydrol-earth-syst-sci.net/13/395/2009/hess-13-395-2009.pdf

Index concept by Gupta et al. (2008) which defines the similarity between data items in a meaningful and hydrologically interpretable way. Instead of working directly with the model output time series, these are converted to vectors of five Signature Index values each, using the Signature Indices proposed by Yilmaz et al. (2008). Once the training is accomplished, every neuron of the map represents the Signature pattern of the model realizations that have been projected to it in the course of the training. It is demonstrated that, this way, the map becomes interpretable in terms of different hydrological characteristics and thereby illustrates the trade-offs in model behaviour. Similar to the previous approach (Appendix A) this mapping is capable of providing preliminary insights into model parameter sensitivities. In addition, however, it is shown that it also allows interpreting the parameter function in the hydrological context, even in the absence of prior knowledge on the model structure. Also, the set of model realizations that most closely approximate the observed discharge time series (in terms of Signature Index pattern) compares favourably to the model realization obtained by minimizing the RMSE using the SCE-UA algorithm. Thus, the results of the second study (Appendix B) indicate that relevant parts of the information contained in the simulated time series can be extracted and reproduced in an efficient and interpretable way by linking the Signature Index concept to the SOM. In addition, by reducing the amount of data processed in the SOM training, the computational burden has been shrunk immensely.

The third contribution (Appendix C[7]) presents a comparative analysis of the distributed hydrological watershed models NASIM (Hydrotec, 2005), LARSIM (Bremicker, 2000) and WaSIM-ETH (Schulla and Jasper, 2001), following an approach similar to the second study (Appendix B). Using five specifically conceptualized indices that extract data on peak discharge characteristics it is examined in how far the individual models are capable of reproducing flood events present in the observed discharge time series. Again, the set of model realizations that most closely approximate the observed discharge time series is determined and compared to reference simulations obtained by implementing a simple calibration strategy by means of the SCE-UA algorithm. The results obtained for NASIM and WaSIM-ETH represent the characteristics of the observed time series with similar or partially superior accuracy, compared to the reference simulations, whereas LARSIM is clearly outperformed. However, it is shown that the poor results for LARSIM are attributable to an obvious incompatibility of LARSIM with respect to the constraints imposed by applying the

---

[7] Herbst, M., Casper, M., Grundmann, J., and Buchholz, O.: Comparative analysis of model behaviour for flood prediction purposes using Self-Organizing Maps, Natural Hazards in Earth System Sciences, 9, 373-392, 2009. http://www.nat-hazards-earth-syst-sci.net/9/373/2009/nhess-9-373-2009.pdf

five index measures. The relative differences between the individual model performances essentially remain intact when the selected model realizations are applied to a validation event with extreme peak discharge. However, in this case, none of the selected models proves capable of properly representing it.

In particular, this study takes advantage of the fact that the SOM allows for a simultaneous evaluation of the data obtained from the three models. The results highlight that each of the models has a very specific behaviour with respect to the representation of peak discharges such that only very few realizations can be characterized as essentially equivalent in terms of the indices. It is argued that the possibility of a direct comparison of model behaviour properties lends itself for potential applications, e.g. in a model ensemble framework where realizations from different models are selected according to the specific requirements of the modelling situation.

4   Discussion, conclusions and further directions

The abovementioned results demonstrate the utility and versatility of SOM for model evaluation purposes. In summary, they indicate:

- advantages of a SOM-based evaluation of Monte-Carlo simulation results in comparison to common methods such as scatterplots and objective function response surfaces, in particular with a view to identifying and understanding parameter sensitivities (see Appendix A and B);

- the possibility of using SOM to identify the set of model realizations (among the Monte-Carlo data) that most closely approximate the observed discharge time series in terms of hydrologically relevant characteristics which, however, can be tailored to fit the requirements of the particular application (Appendix B);

- that SOM can be used for a direct, and relatively objective, comparison of the performances of different model structures (Appendix C).

Based on what is already known about the properties of SOM, it can thus be concluded that these findings corroborate the underlying hypothesis of this work (Sect. 1.5).

The studies presented here, however, only constitute a first step towards using SOM in the context of model evaluation. Nevertheless, the results of the studies as well as the achievements of other researchers encourage a more systematic exploration of its potential in this field. In this respect, SOM offers many aspects and interesting offshoot directions that

deserve further investigation. In principle, there are two promising fields for future efforts: On one hand, the properties of SOM generated mappings and their relationship to more established methods is not yet fully understood. More insights into similarities and differences of the techniques are needed in order to bring about systematic developments of new approaches in conjunction with SOM. On the other hand, the current literature on the application of SOM in the context of environmental modelling does not take into account the more recent developments of self-organizing neural architectures. Thus, given the promising results that have been reported with the 'classic' SOM, there is probably an interesting potential for future developments.

4.1 Scope for further research on the classic SOM

With respect to the 'classic' SOM, a systematic investigation towards gaining a better comprehension of SOM should start with a careful examination of the effects of the data normalization which is carried out prior to the training (see Appendix B, Sect. 2.4): The normalization of simulated data to values having zero mean and a variance of one is very likely to have immense effects on model evaluation because it causes every time step of a time series to have equal importance when used in combination with e.g. least squares statistics (pers. comm. Hoshin V. Gupta, 30.4.2008). Finding more answers on this issue will thus shed new light on the first study (Appendix A). Interestingly, the transformation of discharge time series prior to the evaluation has a parallel in the λ-transformation used in conjunction with the heteroscedastic maximum likelihood error (HMLE; Sorooshian and Dracup, 1980). The HMLE assumes the variance to vary with the magnitude of flow and the λ-transformation is used herein to stabilize the residual variance across the ranges of flow (Yapo et al., 1996). With regard to the SOM training itself, the role of the learning rate parameter, the neighbourhood function and the similarity measure (see Appendix B, Sect. 2.3) should be examined in more detail so as to improve the properties of the SOM for a specific application. Further, it needs to be elucidated to what extent and under which circumstances the SOM and identification of the best-matching unit can be used to assist model uncertainty estimation or optimization strategies. To this end, the statistical properties of SOM should come into play. This issue hitherto has found only marginal attention in this context. Kim et al. (2006) among others, however, already demonstrated that probability densities and likelihoods can be estimated from the SOM. Another important point is that the Signature Indices used in Appendix B (and also those indices used in Appendix C) by no means make a claim to completely cover all relevant information contained in the data. This leads to a

consideration of model in- and output data in the context of information theory: Very little is known concerning the requirements of a 'parsimonious' description of model behaviour, i.e. how to exclude redundant information on model behaviour. Transferred to the SOM context this means that further research on the development of SOM-based applications in the field of hydrological modelling will inevitably be confronted with the question of what constitutes a sufficient and informative metric (how many 'indices' – generally speaking – are required to extract a non-redundant 'fingerprint' of a simulated time series?) and the question concerning the number of neurons required to capture all characteristics of model behaviour. This again leads to the more general problems of time series data mining: Defining and identifying the information content of the data constitutes the key towards a more meaningful and precise application of SOM and all other frameworks for the evaluation and analysis of hydrological models.

Some further motivation to investigate SOM-based local modelling may arise from the field of engineering science, where Principe et al. (1998) present an interesting approach, quite similar to the one introduced by Hsu et al. (2002): They perform non-linear time series modelling on a SOM which is used to represent the reconstructed state space of the observed system. To this end, the phase-space of the input signal is reconstructed by means of delay embedding. Following Takens' theorem (Takens, 1981; Sauer et al., 1991) it is possible to reconstruct the state space of a chaotic dynamical system from a sequence of observations by means of delay embedding, i.e. a sequence of state vectors is created from a given time series $x(t)$, where the current state is $\boldsymbol{x}(t) = [x(t - (N - 1)\tau), x(t - (N - 2)\tau),…,x(t)]^T$, given that the delay τ and N is known. Subsequently, the SOM is trained to form a discretized representation of the input state space and fitting of local linear models is performed directly over a matched neighbourhood of the SOM neurons. The difficulty of phase-space embedding lies in determining the optimal embedding parameters τ and N, i.e. the delay time and the embedding dimension. Additionally, in order that method be effective it is required that the observed system displays chaotic behaviour. Interestingly, Sivakumar et al. (2000) provides "preliminary evidence" (Sivakumar et. al., 2000, p. 407) of the existence of chaos in monthly rainfall-runoff data from the Swedish Göta basin and successfully performed non-linear prediction using the concept of phase-space embedding for reconstructing the attractor of hydrological time series, however, without employing the effective SOM-based technique to represent the state space. In order to perform non-linear prediction also Walter et al. (1990) propose the use of SOM in conjunction with a discretization of the state space. Thus, there is some evidence that may encourage further work in this field.

4.2 Developments in self-organizing neural architectures

So far, all studies (known to the author of this thesis) in the context of SOM and modelling have been based on the application of the 'classic' SOM architecture (or derivatives of the same) which has already been proposed in the early 1980s (Kohonen, 2001). Since the 1990s, however, a series of interesting developments in the field of self-organizing neural algorithms have been reported which may as well give rise to improved applications and novel perspectives in environmental modelling. On one hand, a tendency towards generalizing the SOM can be noted, with a view to expanding its fields of application: With the Adaptive-Subspace SOM (ASSOM) Kohonen (1995) himself introduced a new SOM architecture in which each neuron, instead of a reference vector, represents a subspace. He demonstrated that the ASSOM automatically forms wavelet- and Gabor-type filters when trained with displaced or moving input patterns and thus lends itself for invariant feature detection. An extreme case of generalization can be seen in the development of modular network SOMs (mnSOM; Tokunaga et al., 2005) in which the reference vectors are replaced by functional modules which can even contain another SOM (SOM²; Furukawa, 2005). On the other hand, efforts have been directed towards loosening the constraints of a rigid, two-dimensional grid topology in which the number of neurons is considered to be constant. An exhaustive review of the confusing wealth of publications available on these SOM developments and their offspring is far beyond the scope of this work. In the context of the abovementioned studies and of the present work, however, it is worth mentioning some developments, in particular those which make use of growing neural networks. These are unsupervised or supervised learning networks in which the number of neurons is not pre-defined but evolves during the training process, forming a multidimensional graph-like representation of the data space. This architecture was dubbed Growing Neural Gas (GNG; Fritzke, 1994). In contrast to the related Neural Gas algorithm presented by Martinetz (1993), which is a vector quantization method where only the neighbourhoods are adaptively defined during the training, a GNG is developing additional nodes and connections in the course of the training process and is therefore able to learn topological relations of a given training data set (Fritzke, 1995b). A striking parallel to the aims of Hsu et al. (2002) or perhaps Schütze et al. (2005) can be seen in the work of Fritzke (1995a): Following the basic ideas of GNG he proposes an incremental network model for supervised learning where each unit is associated to a local linear mapping. The error information obtained during the training is used in order to determine where to insert new neurons, whereas the local linear models are interpolated from the neighbours. The author has demonstrated that the so-called GNG-LLM network can be applied to perform

function approximations. By introducing a criterion for the removal of "useless" neurons Fritzke (1997) developed a self-organizing network algorithm which is even capable of following non-stationary distributions.

The fields of application for GNG (and related methods) are basically the same as for the classic SOM. The primary advantage of incremental neural networks, and specifically of GNG, is that, in contrast to SOM, the number of neurons, which is proportional to the preserved amount of potential information, and the dimensionality of the target space is not required to be defined a priori. As the "true" information content of the data which is processed, e.g. in time series modelling, remains unclear it is always desirable to obtain a non-redundant representation of the training data distribution. Hence, GNG based methods may provide some advantages as they produce small and well-generalizing networks (Fritzke, 1994). A self-organizing network algorithm with the capability to follow non-stationary distributions might lend itself for interesting applications in recursive model uncertainty estimation strategies where successively new information has to be processed.

References

Abrahart, R. J., and See, L.: Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments, Hydrological Processes, 14, 2157-2172, 2000.

Abramowitz, G., Gupta, H. V., Pitman, A., Wang, Y., Leuning, R., Cleugh, H., and Hsu, K.-l.: Neural Error Regression Diagnosis (NERD): A Tool for Model Bias Identification and Prognostic Data Assimilation, Journal of Hydrometeorology, 7, 160-177, DOI:10.1175/JHM479.1, 2006.

Abramowitz, G., Pitman, A., Gupta, H. V., Kowalczyk, E., and Wang, Y.: Systematic Bias in Land Surface Models, Journal of Hydrometeorology, 8, 989-1001, DOI:10.1175/JHM628.1, 2007.

Abramowitz, G., and Gupta, H. V.: Toward a model space and model independence metric, Geophysical Research Letters, 35, L05705, doi:10.1029/2007GL032834, 2008.

Abramowitz, G., Leuning, R., Clark, M., and Pitman, A.: Evaluating the Performance of Land Surface Models, Journal of Climate, 21, 5468-5481, DOI:10.1175/2008JCLI2378.1, 2008.

Alhoniemi, E., Hollmén, J., Simula, O., and Vesanto, J.: Process Monitoring and Modeling using the Self-Organizing Map, Integrated Computer Aided Engineering, 6, 3-14, 1999.

Allinson, N. M., and Yin, H.: Self-Organising Maps for Pattern Recognition, in, edited by: Oja, E., and Kaski, S., Elsevier, Amsterdam, 111-120, 1999.

Ambroise, B., Perrin, J. L., and Reutenauer, D.: Multicriterion validation of a semidistributed conceptual model of the water cycle in the Fecht Catchment (Vosges Massif, France), Water Resources Research, 31, 1467-1482, 1995.

Beck, M. B.: The Selection of Structure in Models of Environmental Systems, The Statistician, 35, 151-161, 1986.

Bertalanffy, L. v.: General System Theory, revised edition ed., New York, 295 pp., 1969.

Beven, K.: Changing ideas in hydrology — The case of physically-based models, Journal of Hydrology, 105, 157-172, DOI:10.1016/0022-1694(89)90101-7, 1989.

Beven, K. J., and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrological Processes, 6, 279-298, doi:10.1002/hyp.3360060305, 1992.

Beven, K. J.: Uniqueness of place and process representations in hydrological modelling, Hydrology and Earth System Sciences, 4, 203-214, 2000.

Beven, K. J., and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology, 249, 11-29, doi: 10.1016/S0022-1694(01)00421-8, 2001.

Boogaard, H. F. P. v. d., Mynett, A. E., and Ali, M. S.: Self organizing feature maps for the analysis of hydrological and ecological data sets, in: Hydroinformatics '98, edited by: Babovic, V. M., and Larsen, L. C., Rotterdam, 733-740, 1998.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, Water Resources Research, 36, 3663-3674, 2000.

Bremicker, M.: Das Wasserhaushaltsmodell LARSIM. Modellgrundlagen und Anwendungsbeispiele, Freiburger Schriften zur Hydrologie, 11, 2000.

Brown, J. D., and Heuvelink, G. B. M.: Assessing Uncertainty Propagation Through Physically based Models of Soil Water Flow and Solute Transport, in: Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G., Wiley, 1181-1196, 2005.

Bulygina, N., and Gupta, H. V.: Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, Water Resources Research, 45, W00B13, doi:10.1029/2007WR006749, 2009.

Chang, D.-H.: Analysis and modeling of space-time organization of remotely sensed soil moisture, Ph. D., Department of Civil and Environmental Engineering, University of Cincinnati, Cincinnati, Ohio, 169 pp., 2001.

Clemen, T.: Zur Wavelet-gestützten Validierung von Simulationsmodellen in der Ökologie, Dissertation, Institut für Informatik und Praktische Mathematik, Christian-Albrechts-Universität, Kiel, 111 pp., 1999.

Dooge, J. C. I.: Bringing it all together, Hydrology and Earth System Sciences, 9, 3-14, SRef-ID: 1607-7938/hess/2005-9-3, 2005.

Duan, Q., Sorooshian, S., and Gupta, V. K.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resources Research, 28, 1015-1031, doi:10.1029/91WR02985, 1992.

Duan, Q., Gupta, V. K., and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, Journal of Optimization Theory and Applications, 76, 501-521, doi:10.1007/BF00939380, 1993.

Duan, Q.: Global optimization for watershed model calibration, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application Volume 6, AGU, Washington D.C., 89-104, 2003.

Franchini, M., and Pacciani, M.: Comparative analysis of several conceptual rainfall-runoff models, Journal of Hydrology, 122, 161-219, doi:10.1016/0022-1694(91)90178-K, 1991.

Franks, S. W., Gineste, P., Beven, K. J., and Merot, P.: On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process, Water Resources Research, 34, 787-797, doi:10.1029/97WR03041, 1998.

Fritzke, B.: Fast Learning with Incremental RBF Networks, Neural Processing Letters, 1, 2-5, DOI: 10.1007/BF02312392, 1994.

Fritzke, B.: Incremental Learning of Local Linear Mappings, in: Proceedings of the International Conference on Artificial Neural Networks, ICANN'95: International Conference on Artificial Neural Networks, Paris, 1995, 217-222, 1995a.

Fritzke, B.: A Growing Neural Gas Network Learns Topologies, Advances in Neural Information Processing Systems 7, 625-632, 1995b.

Fritzke, B.: A Self-Organizing Network that Can Follow Non-stationary Distributions, in: ICANN '97: Proceedings of the 7th International Conference on Artificial Neural Networks, ICANN '97: International Conference on Artificial Neural Networks, 1997, 613 - 618, 1997.

Furukawa, T.: SOM of SOMs: Self-organizing Map Which Maps a Group of Self-organizing Maps, in: Artificial Neural Networks: Biological Inspirations – ICANN 2005, Lecture Notes in Computer Science, 391-396, 2005.

Gallart, F., Latron, J., Llorens, P., and Beven, K.: Using internal catchment information to reduce the uncertainty of discharge and baseflow predictions, Advances in Water Resources, 30, 808-823, doi:10.1016/j.advwatres.2006.06.005, 2007.

Gan, T. Y., and Biftu, G. F.: Effects of model complexity and structure, parameter interactions and data on watershed modelling, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application Volume 6, AGU, Washington D.C., 317-329, 2003.

Govindaraju, R. S., and Zhang, B.: Radial-Basis Function Networks, in: Artificial Neural Networks in Hydrology, edited by: Govindaraju, R. S., and Ramachandra Rao, A., Water Science and Technology Library, Kluwer Academic Press, Dordrecht, 93-109, 2000.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resources Research, 34, 751-764, 1998.

Gupta, H. V., Sorooshian, S., Hogue, T. S., and Boyle, D. P.: Advances in Automatic Calibration of Watershed Models, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application Volume 6, AGU, Washington D.C., 9-28, 2003.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, 22, 3802-3813, DOI:10.1002/hyp.6989, 2008.

Gutiérrez, J. M., Cano, R., Cofiño, A. S., and Sordo, C.: Analysis and downscaling multi-model seasonal forecasts in Peru using self-organizing maps, Tellus A, 57, 435-447, DOI: 10.1111/j.1600-0870.2005.00128.x, 2005.

Hall, M. J.: How well does your model fit the data? Journal of Hydroinformatics, 3, 49-55, 2001.

Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, Biometrika, 57, 97-109, doi:10.1093/biomet/57.1.97, 1970.

Hauhs, M., and Lange, H.: Foundations for the simulation of ecosystems, in: Simulation: Pragmatic Constructions of Reality, edited by: Lenhard, J., Küppers, G., and Shinn, T., Sociology of the Sciences, Springer, 57-77, 2006.

Hauhs, M., and Lange, H.: Classification of runoff in headwater catchments: A physical problem? Geography Compass, 2, 235-254, 2008.

Haykin, S.: Neural networks - a comprehensive foundation, 2nd ed., New Jersey, 842 pp., 1999.

Hewitson, B., and Crane, R.: Self-organizing maps: applications to synoptic climatology, Climate Research, 22, 13-26, 2002.

Hsu, K.-l., Gupta, H. V., Gao, X., Sorooshian, S., and Imam, B.: Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis, Water Resources Research, 38, 1302, doi:10.1029/2001WR000795, 2002.

Huang, M., Liang, X., and Liang, Y.: A transferability study of model parameters for the variable infiltration capacity land surface scheme, Journal of Geophysical Research, 108, 8864, doi:10.1029/2003JD003676, 2003.

Hydrotec: Rainfall-Runoff-Model NASIM - program documentation (in German), Hydrotec GmbH, Aachen, 579, 2005.

Islam, S., and Kothari, R.: Spatial organization and characterization of soil physical properties using Self-Organizing Maps, in: Artificial Neural Networks in Hydrology, edited by: Govindaraju, R. S., and Ramachandra Rao, A., Water Science and Technology Library, Kluwer Academic Press, Dordrecht, 199-207, 2000.

Jain, A., and Srinivasulu, S.: Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques, Journal of Hydrology, 317, 291-306, doi:10.1016/j.jhydrol.2005.05.022, 2006.

Kalteh, A. M., and Berndtsson, R.: Interpolating monthly precipitation by self-organizing map (SOM) and multilayer perceptron (MLP), Hydrological Sciences Journal, 52, 305-317, 2007.

Kalteh, A. M., Hjorth, P., and Berndtsson, R.: Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application, Environmental Modelling & Software, 23, 835-845, DOI:10.1016/j.envsoft.2007.10.001, 2008.

Kaski, S.: Data Exploration Using Self Organizing Maps, Dr., Department of Computer Science and Engineering, Helsinki University of Technology, Helsinki, 57 pp., 1997.

Kavetski, D., Franks, S. W., and Kuczera, G.: Confronting Input Uncertainty in Environmental Modelling, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Sciences and Applications Series Volume 6, AGU, Washington D.C., 49-68, 2003.

Kim, T., Eom, I., and Kim, Y.: Texture Segmentation Using SOM and Multi-scale Bayesian Estimation, in: Advances in Neural Networks - ISNN 2006, 661-668, 2006.

Klemeš, V.: Conceptualization and scale in hydrology, Journal of Hydrology, 65, 1-23, 1983.

Kohonen, T.: The adaptive-subspace SOM (ASSOM) and its use for the implementation of invariant feature detection, International Conference on ArtiÞcial Neural Networks, Paris, 1995, 3-10, 1995.

Kohonen, T.: Self-Organizing Maps, 3rd ed., Information Sciences, Berlin, Heidelberg, New York, 501 pp., 2001.

Kuczera, G., and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, Journal of Hydrology, 211, 69-85, doi:10.1016/S0022-1694(98)00198-X, 1998.

Labat, D., Ababou, R., and Mangin, A.: Rainfall-runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses, Journal of Hydrology, 238, 149-178, 2000.

Lamb, R., Beven, K., and Myrabø, S.: Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model, Advances in Water Resources, 22, 305, 1998.

Lane, S. N.: Assessment of rainfall-runoff models based upon wavelet analysis, Hydrological Processes, 21, 586-607, 2007.

Legates, D. R., and McCabe Jr., G. J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resources Research, 35, 233-241, doi:10.1029/1998WR900018, 1999.

Lin, G.-F., and Chen, L.-H.: Identification of homogeneous regions for regional frequency analysis using the self-organizing map, Journal of Hydrology, 324, 1-9, 2006.

Lin, Z., and Beck, M. B.: On the identification of model structure in hydrological and environmental systems, Water Resources Research, 43, W02402, doi:10.1029/2005WR004796, 2007.

Lischeid, G.: Investigating the suitability of self-organizing maps to analyze high-dimensional water quality data sets, Geophysical Research Abstracts, 5, 2003.

Lischeid, G.: A decision support system for mountain basin management using sparse data, Geophysical Research Abstracts, 8, 2006.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Advances in Water Resources, 26, 205-216, 2003.

Martinetz, T. M., Berkovich, S. G., and Schulten, K. J.: "Neural-gas" network for vector quantization and its application to time series prediction, IEEE Transactions on Neural Networks, 4, 558-569, 1993.

Melching, C. S.: Reliability Estimation, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resources Publications, Highlands Ranch, 69-118, 1995.

Mele, P. M., and Crowley, D. E.: Application of self-organizing maps for assessing soil biological quality, Agriculture, Ecosystems & Environment, 126, 139-152, 2008.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of State Calculations by Fast Computing Machines, The Journal of Chemical Physics, 21, 1087-1092, 1953.

Minns, A. W., and Hall, M. J.: Artificial Neural Network Concepts in Hydrology, in: Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G., John Wiley & Sons, Ltd., 307-320, 2005.

Misirli, F., Gupta, H. V., Sorooshian, S., and Thiemann, M.: Bayesian Recursive Estimation of Parameter and Output Uncertainty for Watershed Models, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application Volume 6, AGU, Washington D.C., 113-124, 2003.

Montanari, A., and Toth, E.: Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins? Water Resources Research, 43, doi:10.1029/2006WR005184, 2007.

Moradkhani, H., Hsu, K.-l., Gupta, H. V., and Sorooshian, S.: Improved streamflow forecasting using self-organizing radial basis function artificial neural networks, Journal of Hydrology, 295, 246-262, doi:10.1016/j.jhydrol.2004.03.027, 2004.

Moradkhani, H., Hsu, K.-L., Gupta, H., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resources Research, 41, W05012, doi:10.1029/2004WR003604, 2005.

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

Obayashi, S., and Sasaki, D.: Visualization and Data Mining of Pareto Solutions Using Self-Organizing Map, in: Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science, 71-86, 2003.

Oja, M., Kaski, S., and Kohonen, T.: Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum, Neural Computing Surveys, 3, 1-156, 2003.

Oreskes, N., and Belitz, K.: Philosophical Issues in Model Assessment, in: Model Validation: Perspectives in Hydrological Science, edited by: Anderson, M. G., and Bates, P. D., John Wiley and Sons, London, 23-41, 2001.

Peeters, L., Bação, F., Lobo, V., and Dassargues, A.: Exploratory data analysis and clustering of multivariate spatial hydrogeological data by means of GEO3DSOM, a variant of Kohonen's Self-Organizing Map, Hydrology and Earth System Sciences, 11, 1309-1321, 2007.

Principe, J. C., Wang, L., and Motter, M. A.: Local dynamic modeling with self-organizing maps and applications to nonlinear system identification and control, Proceedings of the IEEE, 86, 2240-2258, doi:10.1109/5.726789, 1998.

Reusser, D. E., Blume, T., Schaefli, B., and Zehe, E.: Analysing the temporal dynamics of model performance for hydrological models, Hydrology and Earth System Sciences Discussions, 5, 3169-3211, 2008.

Sang, H., Gelfand, A. E., Lennard, C., Hegerl, G., and Hewitson, B.: Interpreting self-organizing maps through space-time data models, Annals of Applied Statistics, 2, 1194-1216, 2008.

Sauer, T., Yorke, J. A., and Casdagli, M.: Embedology, Journal of Statistical Physics, 65, 579-616, DOI:10.1007/BF01053745, 1991.

Savenije, H. H. G.: HESS Opinions "The art of hydrology", Hydrology and Earth System Sciences, 13, 157-161, 2009.

Schaefli, B., and Gupta, H. V.: Do Nash values have value? Hydrological Processes, 21, 2075-2080, 2007.

Schaefli, B., and Zehe, E.: Analyzing hydrological model performance in the wavelet spectral domain, Geophysical Research Abstracts, 9, 2, SRef-ID: 1607-7962/gra/EGU2007-A-08667, 2007.

Schoups, G., van de Giesen, N. C., and Savenije, H. H. G.: Model complexity control for hydrologic prediction, Water Resources Research, 44, W00B03, doi:10.1029/2008WR006836, 2008.

Schulla, J., and Jasper, K.: Model description WaSIM-ETH, Zürich, 2001.

Schütze, N., Schmitz, G. H., and Petersohn, U.: Self-organizing maps with multiple input-output option for modeling the Richards equation and its inverse solution, Water Resources Research, 41, W03022, doi:10.1029/2004WR003630, 2005.

Seibert, J., and McDonnell, J. J.: The Quest for an Improved Dialog Between Modeler and Experimentalist, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application Volume 6, AGU, Washington D.C., 301-315, 2003.

Simon, G., Lendasse, A., Cottrell, M., Fort, J. C., and Verleysen, M.: Time series forecasting: Obtaining long term trends with self-organizing maps, Pattern Recognition Letters, 26, 1795-1808, doi:10.1016/j.patrec.2005.03.002, 2005.

Simula, O., Vesanto, J., Alhoniemi, E., and Hollmén, J.: Analysis and Modeling of Complex Systems Using the Self-Organizing Map, in: Neuro-Fuzzy Techniques for Intelligent Information Systems, edited by: Kasabov, N., and Kozma, R., Physica Verlag (Springer Verlag), 3-22, 1999.

Sivakumar, B., Berndtsson, R., Olsson, J., Jinno, K., and Kawamura, A.: Dynamics of monthly rainfall-runoff process at the Gota basin: A search for chaos, Hydrology and Earth System Sciences, 4, 407-417, 2000.

Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, in: Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G., Wiley, 193-220, 2005.

Sorooshian, S., and Dracup, J. A.: Stochastic Parameter Estimation Procedures for Hydrologie Rainfall-Runoff Models: Correlated and Heteroscedastic Error Cases, Water Resources Research, 16, 430-442, 1980.

Sorooshian, S., and Gupta, V. K.: Model calibration, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Colorado, 23-68, 1995.

Spear, R. C., and Hornberger, G. M.: Eutrophication in peel inlet--II. Identification of critical uncertainties via generalized sensitivity analysis, Water Research, 14, 43-49, 1980.

Takens, F.: Detecting strange attractors in turbulence, in: Dynamical Systems and turbulence, edited by: Rand, D. A., and Young, L.-S., Springer Lecture Notes in Mathematics 898, Springer-Verlag, New York, 365-381, 1980.

Thiemann, M., Trosser, M., Gupta, H., and Sorooshian, S.: Bayesian recursive parameter estimation for hydrologic models, Water Resources Research, 37, 2521-2536, 2001.

Tokunaga, K., Furukawa, T., and Yasui, S.: Modular Network SOM: Self-Organizing Maps in Function Space, Neural Information Processing - Letters and Reviews, 9, 15-22, 2005.

Torrence, C., and Compo, G. G.: A Practical Guide to Wavelet Analysis, Bulletin of the American Meteorological Society, 79, 61-78, 1998.

Troch, P. A., Carrillo, G. A., Heidbüchel, I., Rajagopal, S., Switanek, M., Volkmann, T. H. M., and Yaeger, M.: Dealing with Landscape Heterogeneity in Watershed Hydrology: A Review of Recent Progress toward New Hydrological Theory, Geography Compass, 3, 375-392, 2009.

Vesanto, J.: Using the SOM and Local Models in Time-Series Prediction, Workshop on Self-Organizing Maps (WSOM'97), Espoo, Finland, 1997, 209-214, 1997.

Vesanto, J.: SOM-based data visualization methods, Intelligent Data Analysis, 3, 111-126, DOI:10.1016/S1088-467X(99)00013-X, 1999.

Vesanto, J.: Using SOM in Data Mining, Licentiate's thesis, Department of Computer Science and Engineering, Helsinki University of Technology, Helsinki, 57 pp., 2000.

Vrugt, J. A., Bouten, W., and Weerts, A. H.: Information Content of Data for Identifying Soil Hydraulic Parameters from Outflow Experiments, Soil Science Society of America Journal, 65, 19-27, 2001.

Vrugt, J. A., Bouten, W., Gupta, H. V., and Sorooshian, S.: Toward improved identifiability of hydrologic model parameters: The information content of experimental data, Water Resources Research, 38, doi:10.1029/2001WR001118, 2002.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resources Research, 39, 1214, doi:10.1029/2002WR001746, 2003a.

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resources Research, 39, 1201, doi:10.1029/2002WR001642, 2003b.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, Water Resources Research, 41, W01017, doi:10.1029/2004WR003059, 2005.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., and Gupta, H. V.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis, Hydrological Processes, 17, 455-476, 2003a.

Wagener, T., Wheater, H. S., and Gupta, H. V.: Identification and Evaluation of Watershed Models, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application Volume 6, AGU, Washington D.C., 29-47, 2003b.

Wagener, T., Wheater, H. S., and Gupta, H. V.: Rainfall-Runoff Modelling in Gauged and Ungauged Basins, Imperial College Press, London, 306 pp., 2004.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment Classification and Hydrologic Similarity, Geography Compass, 1, 901-931, 2007.

Walter, J., Ritter, H., and Schulten, K.: Nonlinear prediction with self-organizing maps, IJCNN International Joint Conference on Neural Networks, San Diego, CA, 1990, 589-594, 1990.

Warren Liao, T.: Clustering of time series data--a survey, Pattern Recognition, 38, 1857-1874, doi:10.1016/j.patcog.2005.01.025, 2005.

Weerts, A. H., and El Serafy, G. Y. H.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resources Research, 42, W09403, doi: 10.1029/2005WR004093, 2006.

Willmott, C. J.: On the validation of models, Phys. Geogr., 2, 184-194, 1981.

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data, Journal of Hydrology, 181, 23-48, doi:10.1016/0022-1694(95)02918-4, 1996.

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, Journal of Hydrology, 204, 83-97, doi:10.1016/S0022-1694(97)00107-8, 1998.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resources Research, 44, doi:10.1029/2007WR006716, 2008.

Zehe, E., and Blöschl, G.: Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions, Water Resources Research, 40, W10202, doi:10.1029/2003WR002869, 2004.

Appendix A

**Hydrology and
Earth System
Sciences**

Towards model evaluation and identification using Self-Organizing Maps

M. Herbst and M. C. Casper

Department of Physical Geography, University of Trier, Germany

Abstract. The reduction of information contained in model
time series through the use of aggregating statistical perfor-
mance measures is very high compared to the amount of in-
formation that one would like to draw from it for model iden-
tification and calibration purposes. It has been readily shown
that this loss imposes important limitations on model identifi-
cation and -diagnostics and thus constitutes an element of the
overall model uncertainty. In this contribution we present an
approach using a Self-Organizing Map (SOM) to circumvent
the identifiability problem induced by the low discrimina-
tory power of aggregating performance measures. Instead, a
Self-Organizing Map is used to differentiate the spectrum of
model realizations, obtained from Monte-Carlo simulations
with a distributed conceptual watershed model, based on the
recognition of different patterns in time series. Further, the
SOM is used instead of a classical optimization algorithm
to identify those model realizations among the Monte-Carlo
simulation results that most closely approximate the pattern
of the measured discharge time series. The results are an-
alyzed and compared with the manually calibrated model
as well as with the results of the Shuffled Complex Evo-
lution algorithm (SCE-UA). In our study the latter slightly
outperformed the SOM results. The SOM method, however,
yields a set of equivalent model parameterizations and there-
fore also allows for confining the parameter space to a region
that closely represents a measured data set. This particular
feature renders the SOM potentially useful for future model
identification applications.

1 Introduction

Information from existing or additional observed sources is
crucial to decrease model uncertainty. Model evaluation and
model identification usually resort to aggregating statistical

measures to compare observed and simulated time series
(Legates and McCabe Jr., 1999). In this context these mea-
sures involve considerable problems (Yapo et al., 1998; Lane,
2007): many objective functions imply assumptions about
the error properties that are often violated when dealing with
the agreement between measured and simulated time series:
very often the errors are (1) not normally distributed, (2) do
not have a mean of zero, (3) show autocorrelation or (4) are
heteroscedastic. To certain extent, the choice of performance
measures can be made such that the evaluation makes cer-
tain emphasis on different parts of the hydrograph (Gupta et
al., 1998). Yet aggregating measures of performance have in
common that the information contained in the errors is aggre-
gated into a single numerical value, regardless of the charac-
teristic and the actual pattern of the error. In consequence,
essentially different model results can be obtained with close
to identical performance measure values although the param-
eter sets used to generate them are widely scattered through-
out the parameter space. Therefore the information conveyed
by aggregating goodness-of-fit measures is likely to be insuf-
ficient to unambiguously differentiate between a number of
alternative model realizations (and thus cannot give evidence
of their equivalence, i.e. equifinality) (Gupta et al., 2003;
Beven and Binley, 1992). This lack of discriminatory power
imposes limitations on model identification and constitutes
an important source of model uncertainty (Wagener et al.,
2003). In contrast, one strong point of the manual calibration
procedure resides on the ability to use various complemen-
tary information sources, e.g. river discharge, groundwater
level or soil moisture observations (Franks et al., 1998; Lamb
et al., 1998; Seibert, 2000; Ambroise et al., 1995). In the first
instance, however, its success is due to the simultaneous eval-
uation of numerous different characteristics related to a time
series (Gupta et al., 2003) which allows for a much better
extraction of information from the available data. Multiple-
criteria approaches that seek to emulate this strategy to some
extent have therefore considerably improved model identifi-
ability. Important examples of successful applications of this
strategy can be found e.g. in Gupta et al. (1998), Boyle et

Correspondence to: M. Herbst
(herbstm@uni-trier.de)

al. (2000), Vrugt et al. (2003) and Wagener et al. (2004). Yet model identification methods that depend on common statistical approaches might still not be able to extract enough information relevant to this task (Gupta et al., 2003). An exciting new point of view for model evaluation and identification that tackles these shortcomings emerges from the transformation of the existing data into the frequency domain and the wavelet domain (e.g. Clemen, 1999; Lane, 2007; Montanari and Toth, 2007).

In order to improve model identifiability (as proposed by Gupta et al., 1998; Yapo et al., 1998; Boyle et al., 2000) and the extraction of information from existing data we introduce an approach that, in a sense, emulates the visual assessment of model hydrographs. To circumvent the ambiguity induced by standard objective functions a Self-Organizing Map (SOM) (Kohonen, 2001) is used to represent the spectrum of model realizations obtained from Monte-Carlo simulations with a distributed conceptual watershed model based on the recognition of different patterns of model residual time series.

Self-Organizing maps have found successful practical applications in speech recognition, image analysis, categorization of electric brain signals (Kohonen, 2001) as well as process monitoring (Alhoniemi et al., 1999; Simula et al., 1999) and local time series modelling (Vesanto, 1997; Principe et al., 1998; Cho, 2004). Similarly diverse are the currently emerging applications of SOM in the field of hydrology: Examples for the analysis of hydrochemical data can be found in Peeters et al. (2007) and Lischeid (2006). Schütze et al. (2005) apply a variant of the SOM to approximate the Richards equation and its inverse solution. Hsu et al. (2002) successfully performed system identification and daily streamflow predictions with the Self-Organizing Linear Output Mapping Network (SOLO). They used a SOM to control local regression functions according to the stage of the rainfall-runoff process. Kalteh and Berndtsson (2007) use SOM for the interpolation of monthly precipitation.

In Sect. 2.1 of this contribution we summarize the principles and advantages of SOM and describe how this method is applied to yield a topologically ordered mapping of model output time series according to the similarity in the temporal patterns of their residuals obtained through Monte-Carlo simulations. The properties of this "semantic map" of model realizations will be examined by relating the map elements (i) to the standard performance measures of the associated model runs and (ii) to the parameter values that have been used to generate the model results. It is shown that a SOM is capable of giving visual insights into the parameter sensitivity and the operating of the model structure. Moreover, in the second part of this article these properties are used to introduce an application of the Self-Organizing Map for parameter identification purposes. The SOM is used to identify those model realizations among a given set of Monte-Carlo simulations results that most closely approximate the pattern of the measured time series, i.e. the "zero-residual"

realization. The result will be analyzed and compared with the manually calibrated model as well as with the result of the single-objective Shuffled Complex Evolution algorithm (SCE-UA, Duan et al., 1993).

2 Methods

2.1 The Self-Organizing Map

The Self-Organizing Map is a type of artificial neural network (ANN) and unsupervised learning algorithm that is used for clustering, visualization and abstraction of multi-dimensional data. Unlike other types of ANN it has no output function. Instead it maps vectorial input data items with similar patterns onto contiguous locations of a discrete low-dimensional grid of neurons in a topology-preserving manner. Therefore its output can be compared to a semantic map: nearby locations on the map are attributed similar data patterns. Each of the map's neurons becomes "sensitized" to a different domain of the patterns contained in the vectorial training data items, i.e. the map units act as decoder for different types of patterns contained in the input data (Kohonen, 2001).

Each input data item $x \in X$ is considered as a vector

$$x = [x_1, x_2, \ldots, x_n]^T \in \Re^n \quad (1)$$

with n being the dimension of the input data space. A fixed number of k neurons indexed i is arranged on a regular grid G with each neuron being associated to a weight vector

$$m_i = [\mu_{i1}, \mu_{i2}, \ldots, \mu_{in}]^T \in \Re^n \quad (2)$$

also called reference vector, which has the same dimensionality as the input vectors $x \in X$. These weights connect each input vector x in parallel to all neurons of G. Moreover the neurons are connected to each other. In our case this interconnection is defined on a hexagonal grid topology. The training of the SOM now comprises the following steps (Fig. 1):

1. The components of the m_i are initialized with a sequence of values from points on the plane spanned by the two greatest eigenvectors of the data distribution. This procedure assures a faster and more reliable convergence of the algorithm (Kohonen, 2001).

2. Randomly pick an input vector sample $x \in X$ and compute the Euclidean distance

$$\|x - m_i\| = \sqrt{\sum_{j=1}^{n} (x_j - m_{ij})^2} \quad (3)$$

between x and each of the reference vectors m_i (as a measure of similarity, generally any other metric can be applied as well) and find the neuron $c(x)$ with a reference vector m_c such that

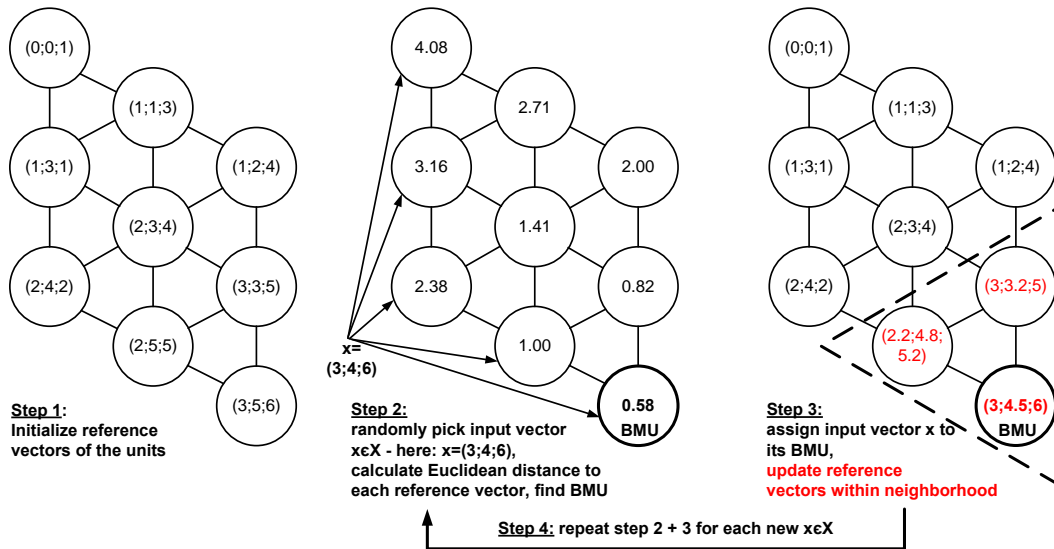$$\|x - m_c\| = \min_i \{\|x - m_i\|\} \quad (4)$$

Fig. 1. The basic steps of the SOM algorithm.

c is then called the best-matching unit (BMU) and defines the image of the sample x on the map G.

3. The nodes that are within a certain distance of the "winning neuron" c are updated according to the equation

$$\boldsymbol{m}_i\,(t+1) = \boldsymbol{m}_i\,(t) + \alpha\,(t)\,h_{ci}\,(t)\,[\boldsymbol{x}\,(t) - \boldsymbol{m}_i\,(t)] \qquad (5)$$

where t is the number of the iteration step and $\boldsymbol{m}_i(t)$ is the current weight vector which is updated proportionally to the difference $[\boldsymbol{x}(t)-\boldsymbol{m}_i(t)]$. $h_{ci}(t)$ determines the degree of neighbourhood between the winning neuron c and neuron i for an input $\boldsymbol{x}\in X$, i.e. the rate of adaptation in the neighbourhood around c. This function is required to be symmetric about c and decreasing to zero with growing lateral distance from c (Haykin, 1999). Commonly the Gaussian function

$$h_{ci}\,(t) = \exp\left(-\frac{\|\boldsymbol{r}_c - \boldsymbol{r}_i\|^2}{2\sigma^2\,(t)}\right) \qquad (6)$$

is used, whereas $\|\boldsymbol{r}_c - \boldsymbol{r}_i\|^2$ denotes the lateral distance between the winning neuron and the neuron i.

$\sigma(t)$ defines the width of the topological neighbourhood, which is also monotonically decreasing with t. It is required that $h_{ci}(t)\to 0$ for $t\to\infty$. In Eq. (5) $\alpha(t)$ is called the learning rate factor ($0<\alpha(t)<1$) which proportionally to the iteration step t monotonically decreases the rate of change of the weight vectors. According to Kohonen (2001) an exact choice of the function is not relevant. With Eq. (5) the training acquires adaptive and cooperative properties through which the weights $\boldsymbol{m}_i$ are updated to move closer towards the winning neuron, similar to an elastic net (Kohonen, 2001).

4. Repeat steps 2 and 3 with the next data vector $\boldsymbol{x}$ until a fixed number of iterations is reached.

Upon repeated cycling through the training data the mapping from the continuous input space X onto the spatially discrete output space G acquires the following properties (Haykin, 1999):

– The reference vectors $\boldsymbol{m}_i$ "follow" the distribution of the input data vectors such that the map G provides a discrete approximation to the input space X. This is as well the reason why dimensionality reduction and data compression properties can be attributed to the SOM. The fix number of weight vectors $\boldsymbol{m}_i$ can be interpreted as pointers for their corresponding neuron into the input space X, hence the elements of $\boldsymbol{m}_i$ can be interpreted as coordinates of the image of this neuron in the input space.

– From Eq. (5) immediately follows the topological ordering property of the mapping computed by the SOM such that the location of a neuron on the grid G represents a particular domain of pattern in the input data. Moreover, this ordering property at the same time provides fault and noise tolerant abilities of the mapping (see also Allinson and Yin, 1999). The local interactions between the neurons provide for the smoothness of the map.

– Patterns in the input space X that occur more frequently are mapped onto a larger area in the output space G.

– A SOM has the ability to select a best set of features for approximating an underlying nonlinear distribution corrupted by additive noise. Hence SOM provides a discrete approximation of principle curves, i.e. a generalization of principal component analysis.

In the second part of this contribution we make use of the fact that the SOM can also be applied to project an input data

www.hydrol-earth-syst-sci.net/12/657/2008/

Hydrol. Earth Syst. Sci., 12, 657–667, 2008

Table 1. Statistical goodness-of-fit measures calculated for the model output (Qobs: observed discharge, Qsim: simulated discharge).

Name	Description	Formula
BIAS	Mean error	$\frac{1}{N}\sum\limits_{k=1}^{N}\left(\text{Qobs}-\text{Qsim}_k\right)$
RMSE	Root of mean squared error	$\sqrt{\frac{1}{N}\sum\limits_{k=1}^{N}\left(\text{Qobs}-\text{Qsim}_k\right)^2}$
CEFFlog	Logarithmized Nash-Sutcliffe coefficient of efficiency	$\frac{\sum\limits_{k=1}^{N}\left(\ln(\text{Qobs})-\ln(\text{Qsim}_k)\right)^2}{\sum\limits_{k=1}^{N}\left(\ln(\text{Qobs})-\ln(\bar{\text{Q}}\text{obs})\right)^2}$
IAg	Willmott's index of agreement	$1-\frac{\sum\limits_{k=1}^{N}\left(\text{Qobs}-\text{Qsim}_k\right)^2}{\sum\limits_{k=1}^{N}\left(\left|\text{Qsim}_k-\bar{\text{Q}}\text{obs}\right|+\left|\text{Qobs}-\bar{\text{Q}}\text{obs}\right|\right)^2}$
	(Willmott, 1981, 1982) $0 \leq \text{IAg} \leq 1$	
MAPE	Mean average percentual error	$\frac{100}{N}\sum\limits_{k=1}^{N}\frac{1}{\text{Qobs}}\left|\text{Qsim}_k-\text{Qobs}\right|$
VarMSE	Variance part of the mean squared error	$\frac{\sqrt{\frac{1}{N}\sum\limits_{k=1}^{N}\left(\text{Qobs}-\bar{\text{Q}}\text{obs}\right)^2}-\sqrt{\frac{1}{N}\sum\limits_{k=1}^{N}\left(\text{Qsim}-\bar{\text{Q}}\text{sim}\right)^2}}{\frac{1}{N}\sum\limits_{k=1}^{N}\left(\text{Qobs}-\text{Qsim}\right)^2}$
Rlin	Coefficient of determination	$\frac{\sum\limits_{k=1}^{N}\left[\left(\text{Qsim}-\bar{\text{Q}}\text{sim}\right)\left(\text{Qobs}-\bar{\text{Q}}\text{obs}\right)\right]}{\sqrt{\sum\limits_{k=1}^{N}\left(\text{Qsim}-\bar{\text{Q}}\text{sim}\right)^2\sum\limits_{k=1}^{N}\left(\text{Qobs}-\bar{\text{Q}}\text{obs}\right)^2}}$

vector y onto the discrete output space that has not been part of the training data manifold. This means that according to Eq. (4) a neuron $c(y)$ with reference vector $m_{c(y)}$ is activated for which

$$\left\|y-m_{c(y)}\right\| = \min_{i}\left\{\left\|y-m_i\right\|\right\} \tag{7}$$

The "image" $c(y)$ of the projected data item y then represents the domain of input data patterns from X that is most similar to y. Moreover, as the number of neurons k is much smaller than the number of vectors used for the training, this neuron will be 'sensitized' and associated to a number of input data patterns from X which will represent the domain of input data patterns that is closest to y.

2.2 Experimental setup

In our example 4000 residual time series (i.e. the element-wise difference between the simulated and the observed time series vectors) constituted the input data vectors of the training data set. The model time series were obtained from 4000 Monte Carlo simulations (see Sect. 2.3) with the distributed conceptual watershed model NASIM running at hourly time steps over a period of two years, i.e. each input data vector consisted of 17 472 elements. Before the training, normalization of the data after Eq. (8) was carried out to avoid that high data values (vector elements) dominate the training because of their higher impact on the Euclidean distance

measure Eq. (3) (Vesanto et al., 2000). Each element of the input data vectors is normalized to a variance of one and zero mean value using the linear transformation

$$x' = (x-\bar{x})/\sigma_x. \tag{8}$$

The dimensions of the SOM were determined using a heuristic algorithm (see Vesanto et al., 2000). A coarse training period of 500 iterations with a large radius for the neighbourhood function was performed followed by a fine tuning period comprising 100 000 training cycles with short neighbourhood radius. In order to compare the results of the aforementioned Monte-Carlo simulation and the properties of the SOM, seven measures of performance, listed in Table 1, were calculated for each model run. Consecutively, a reference data set, which has not been part of the training data, consisting of the time series of observed data was projected onto the SOM according to Sect. 2.1. The resulting time series from this experiment were finally evaluated visually as well as by means of different diagnostic plots. For the model realizations that had been associated with each map element (as a consequence of the training) the means of the corresponding performance measure values were calculated for each measure individually. The map element for which such a mean value is maximum (or minimum according to the measure) is marked as the performance optimum of the corresponding objective function. To determine a common optimum (i.e. balance point) for the seven different performance
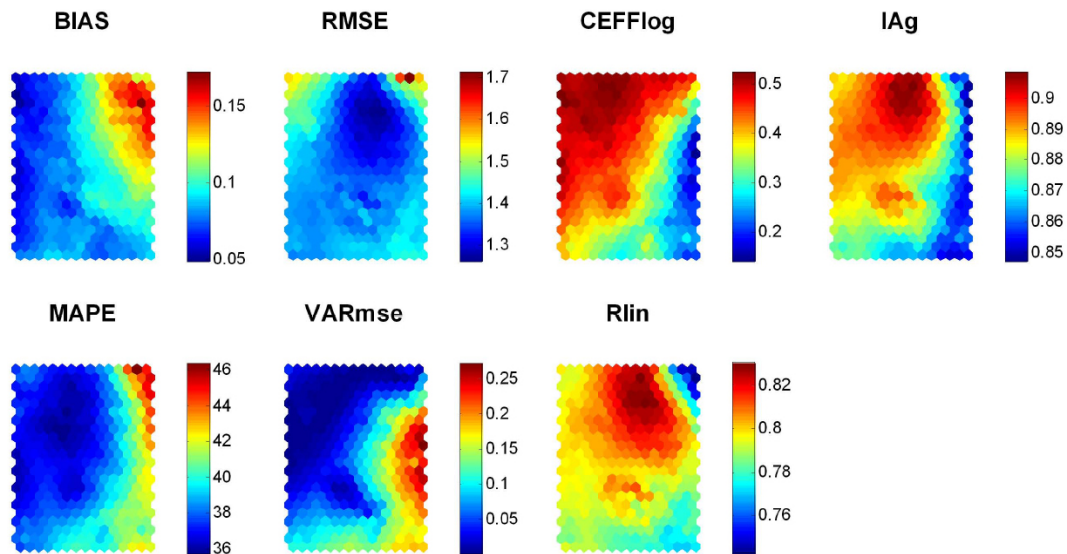
Fig. 2. Distribution of the mean values of each performance measure from Table 2 over the SOM lattice.

optima on the map, each optimum is considered as a mass point with unit weight on the SOM grid. The common optimum is calculated as geometric center of mass resulting from the locations of the seven objective function optima. The x-coordinate x_{opt} of the common optimum for n optima is calculated using Eq. (9)

$$x_{opt} = \frac{\sum w_i x_i}{\sum w_i} \quad , \quad i = 1 \ldots n \tag{9}$$

where x_i are the x-coordinates of the individual optima on the SOM grid. Here $n=7$ and $w_i = 1$ for all i. The y-coordinate y_{opt} is determined accordingly. To ascertain whether the type of data used for the training exerts any influence on the SOM result the experiments were repeated with a SOM trained on discharge time series instead of residual time series. The SCE-UA algorithm (Duan et al., 1993) for the model optimization was run with a maximum of 10 000 iterations and 5 complexes (with 5 points each). For successful termination a change of less than 0.05% of the performance criterion in three consecutive loops was imposed.

2.3 NASIM model and data

NASIM is a distributed conceptual rainfall-runoff model (Hydrotec, 2005). It uses non-linear storage elements to simulate the soil water balance on spatially homogeneous units with respect to soil and land use, which themselves are subdivided into soil layers. NASIM is being commercially distributed since the mid-eighties and since then has found widespread application, e.g. in communal water resources management throughout Germany. The details of the model are beyond the scope of this contribution. Instead, we adopt the decision-maker's point of view and treat the model

Table 2. Free NASIM model parameters of the Monte-Carlo simulation with their respective parameter ranges.

Name	Description	Range
RetBasis	Storage coefficient for baseflow component [h]	0.5–3.5
RetInf	Storage coefficient for interflow component [h]	2.0–6.0
RetOf	Storage coefficient for surface runoff from unsealed surfaces [h]	2.0–6.0
StFFRet	Storage coefficient for surface runoff from urban areas [h]	2.0–6.0
hL	Horizontal hydraulic conductivity factor	2.0–8.0
maxInf	Maximum infiltration factor	0.025–1.025
vL	Vertical hydraulic conductivity factor	0.005–0.105

as a black-box. Seven parameters were selected for Monte-Carlo random sampling (Table 2). A priori knowledge was used to confine the parameter space: the fixed parameters as well as the ranges of the free parameters were chosen to be identical to those that participated in the course of a manual expert calibration for the test watershed. The bounds reproduce the plausible parameter space for this catchment. No pre-imposed correlation between model parameters has been assumed. The input data for the model was taken from the 129 km^2 low-mountain range test watershed "Schwarze Pockau" Saxony (Germany), situated near the border to Czechoslovakia and tributary of the Freiberger Mulde, a sub-basin of the Elbe River. The period from 1 November 1994 to 28 October 1996 with hourly discharge and precipitation measurements was chosen to drive the Monte-Carlo simulation.
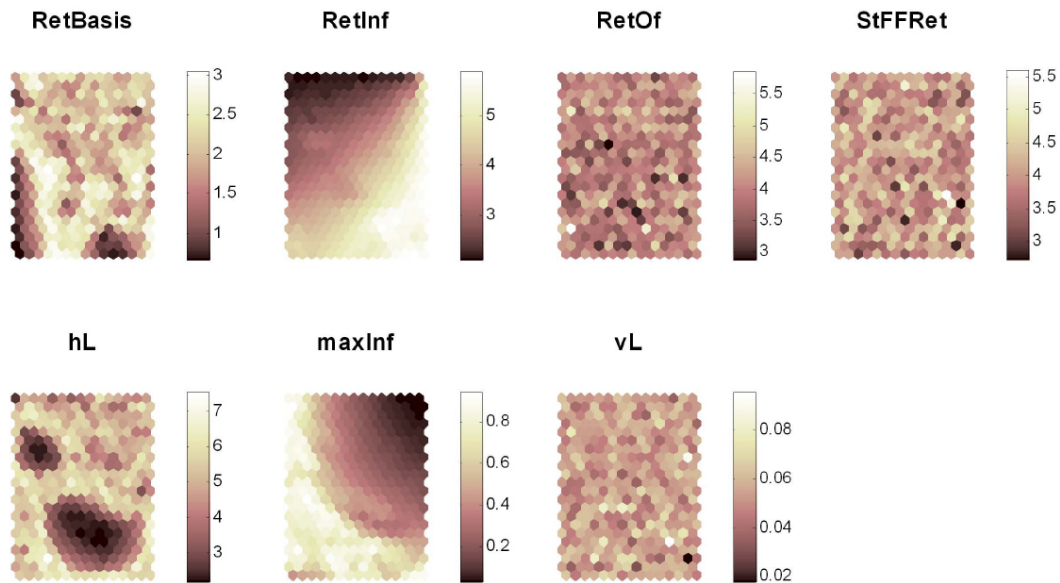
Fig. 3. Distribution of the mean values of each model parameter from Table 1 over the SOM lattice.

3   Results

In the first part of this section the properties of the SOM trained on residual time series and the relation of its elements to the traditional performance measures are examined. The second part is dedicated to testing the projection of measured data onto the SOM.

3.1   Testing the properties of the SOM

After the training each neuron of the 22×15 SOM is expected to be activated by a narrow domain of residual patterns from the input data manifold. The neurons and their respective location on the map are identifiable by index numbers. As the number of neurons is still much smaller than the number of model realizations used for the training, each neuron represents a set of Monte-Carlo model realizations that are characterized through similar temporal patterns with respect to their residuals or discharge values respectively. Because of the topographic ordering principle neighbouring map units, in turn, are expected to be "tuned" to similar residual patterns as well. Because the model realizations used for the training can be referenced by their corresponding index number on the map, the ordering principles of the "semantic map" represented by the SOM can be examined. To this end, the means of different performance measures as well as the mean values of the model parameters on each map element are calculated according to Sect. 2.2. This allows to assess the properties of the map's ordering principle with respect to well known attributes such as (a) the distribution of performance measures and (b) the distribution of different model parameter values over the map lattice. Referring to (a) seven performance measures have been calculated for each model

realization (Table 1). For each of them individual SOM lattices were colour-coded according to the mean of the performance measure of the model runs associated with each map unit. Figure 2 shows the distribution of the performance measures from Table 1 on the SOM lattice. The same procedure was repeated for the values of the free parameters such that the distribution of mean parameter values can be shown for each parameter individually (Fig. 3). In each lattice of Figs. 2 and 3 the positions of the neurons remain identical such that each map element refers to identical model realizations in both figures.

As a striking feature of Fig. 2 it can be seen that, without providing explicit information about the performance measures with the training data, the different performance values are not distributed randomly across the map but significantly relate to different regions of the lattices. To interpret Fig. 2 it is important to notice that warm colours always correspond to high mean values and vice-versa, irrespective of whether this quality is associated with high or low goodness of fit. As to Fig. 3, a visibly ordered relation of the map regions to different parameter values can only be stated for two parameters (RetInf and maxInf), whereas the values of RetOf, StFFRet and vL do not appear to relate to any ordering principle. A similar random pattern can be observed for the two remaining parameters (RetBasis and hL) throughout wide areas of the map. As can be seen from the locally ordered colour distribution, some intercalated areas in these lattices markedly display again a relationship between the parameter values and map locations (which stand for a certain domain of simulated time series pattern). To facilitate the interpretation of these findings we compared Fig. 3 with scatterplots of performance measures. Figure 4 indicates that only the parameters
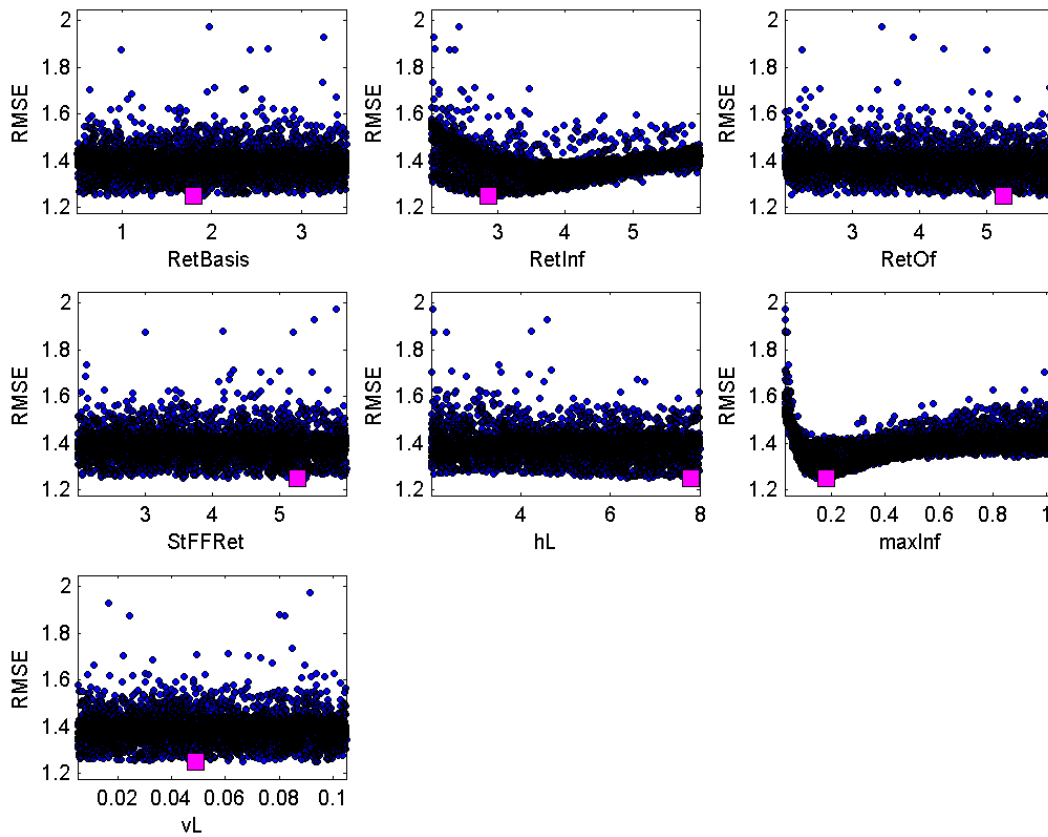
Fig. 4. Scatterplots of RMSE values for each of the examined NASIM parameters.

Table 3. Summary of the parameter values of the 11 model realizations associated to the Best-Matching map Unit when the time series vector of observed discharges is projected onto the SOM.

	RetBasis	RetInf	RetOf	StFFRet	hL	maxInf	vL
min	0.699	4.336	2.379	2.202	2.191	0.107	0.008
max	3.143	4.787	5.731	5.581	6.540	0.134	0.105
mean	1.756	4.555	4.278	3.548	4.674	0.122	0.065

Table 4. Comparison of model performances for results obtained from manual calibration, optimization with SCE-UA and the SOM application. In case of the SOM mean values of 11 results are given.

	BIAS	RMSE	CEFFlog	IAg	MAPE	VARmse	Rlin
manual calibration	0.32	1.58	0.50	0.86	42.36	0.01	0.75
SCE-UA optimization	0.10	1.25	0.49	0.91	36.37	0.06	0.83
SOM (means)	0.13	1.34	0.30	0.88	40.71	0.19	0.81

RetInf and maxInf are sensitive with reference to the RMSE. Scatterplots for the remaining objective functions in Table 1 yielded comparable results. From the aforementioned findings we infer that the locally ordered parameter mean values in Fig. 3 (RetBasis and hL) indicate that the corresponding parameters are subject to interaction with other parameters. Results identical to Figs. 2 and 3 were obtained by training a SOM on discharge time series instead of residual values.

3.2 Projecting the observed time series onto the SOM

To locate the best-matching unit (BMU) of the measured discharge (i.e. zero-residual) time series on the map, according to Sect. 2.1, an input vector consisting of elements with value 0 is constructed. Subsequently, in order to be projected properly along with the training data, the transformation Eq. (8) is carried out using the normalization parameters obtained from the input data set. In Fig. 5 the location of the resulting vector is displayed on top of the performance measure distributions shown in Fig. 2 (black dot). Additionally, the location of the common optimum for the seven performance measures, determined according to Sect. 2.2, is marked (white cross). Figure 6 shows the positions of the different optima on the SOM. It can be seen that the position of the BMU neither coincides with any of the expected objective function optima nor with the common optimum location of the seven performance measures. Additionally,
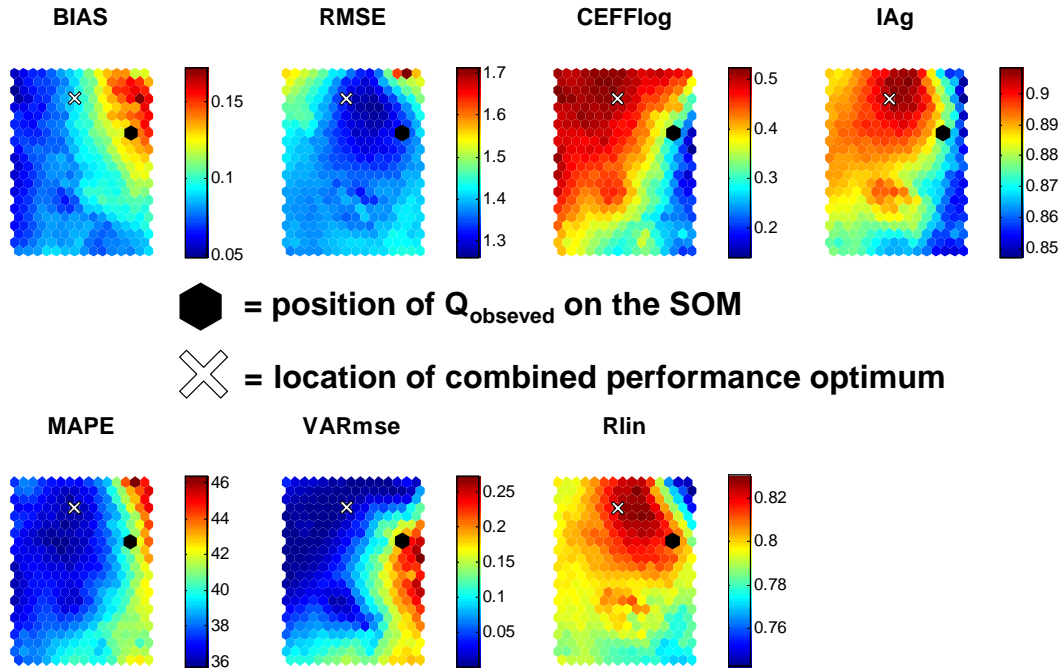
Fig. 5. The location of the best-matching unit (indicated by the black dot) for an input vector that represents the measured discharge time series. The white cross marks the common optimum (balance point) of the seven performance measures on the SOM grid.
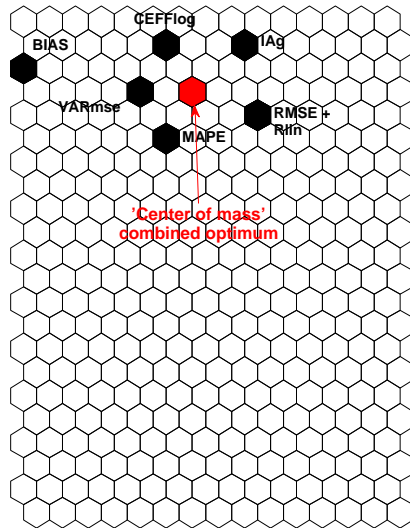

Fig. 6. Determination of the common optimum location (balance point) as the geometric center of mass resulting from the locations of the seven objective function optima on the SOM. Each optimum has unit weight. Here the RMSE and Rlin optima coincide on the same position.

Table 3 summarizes the parameter values of the 11 model realizations that are associated to the BMU for representing the model time series that are most "similar" to a "perfect match" (i.e. the zero-residual case). By comparing these parameters

of the corresponding model runs to the ranges in Table 2 it becomes obvious that, with the exception of RetInf and max-Inf, all parameter values span the full range of the Monte-Carlo sampling bounds. All model realizations attached to this BMU have in common that only the mentioned parameter values appear to be narrowly constrained to values between 4.336 and 4.787 for RetInf and 0.107 and 0.134 for maxInf, respectively. The resulting model outputs for these 11 realizations are shown in Fig. 7c along with the total envelope range of all 4000 simulation outputs in the background and the observed discharge. In order to better point out the differences between the hydrographs only the characteristic time period from 14 January 1995 to 21 October 1995 is reproduced. It can be seen that, compared to the whole set of Monte-Carlo outputs, these realizations obviously comprise a compact subset of "similar" time series. Additionally, the model results obtained from an expert manual calibration and the single-objective automatic calibration using the SCE-UA algorithm (Duan et al., 1993) with the RMSE as objective function (Table 1) are shown in Fig. 7. Although the SOM procedure, unlike the manual calibration, emphasizes all features of the hydrograph equally, the time series associated to the BMU of the measured discharge appear to outperform the result of the expert calibration (Fig. 7a). As expected, the RMSE of the SCE result is smaller than the corresponding values for the BMU results and the manual calibration time series, respectively (Table 4). The same holds true for most of the remaining performance measures such that, in terms of objective function values, the result of the SCE optimization
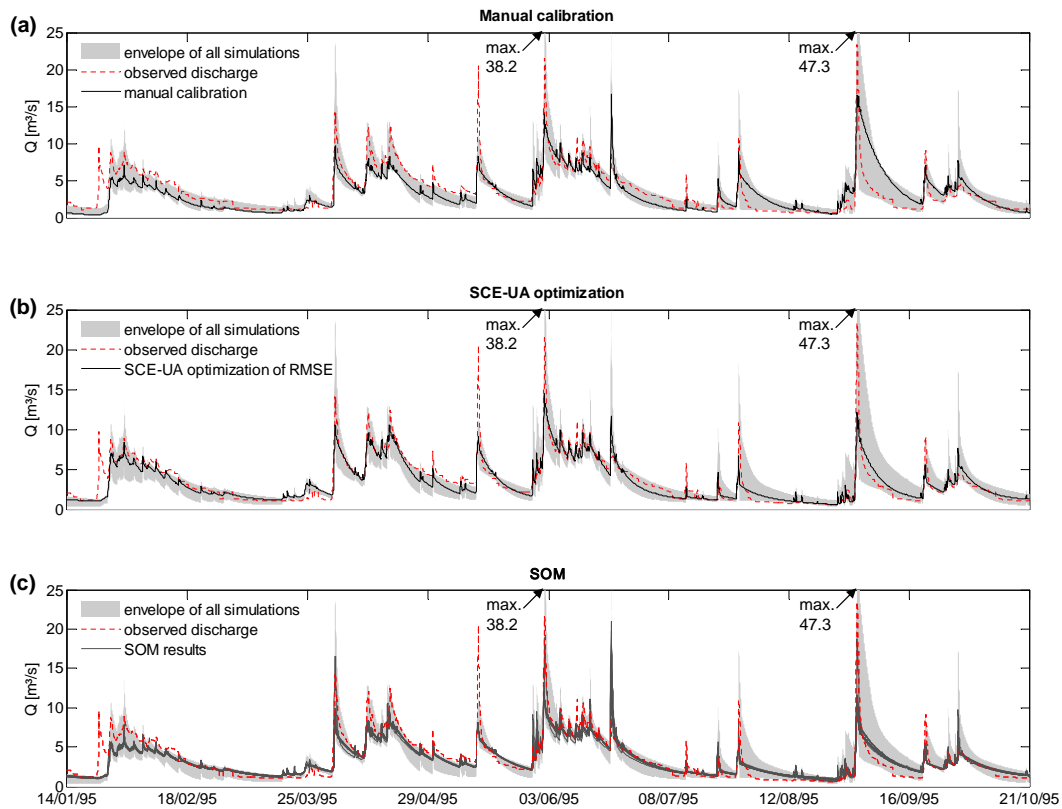
Fig. 7. The model realizations as resulting from **(a)** manual calibration, **(b)** optimization with the SCE-UA algorithm and **(c)** the BMU of the SOM for the measured discharge time series. The time-series are compared to the measured discharge and the envelope of the Monte-Carlo simulation, i.e. the area which is spanned by all model time series for the bounds given in Table 2.

(Fig. 7b) outperforms the BMU realizations as well as the manual calibration. The RMSE of the SCE-optimized model equals the lowest RMSE value obtainable from the given set of Monte-Carlo realizations. The corresponding hydrograph provides a reasonable representation of the measured time series, except for some deficits in the reproduction of the peak discharges. The SCE-optimized model and the realizations from the BMU display noticeable differences in the reproduction of the peak discharges and recession limbs. Though, the SCE result, based on a visual examination, slightly outperforms the SOM method. The training on discharge data time series yields identical results with respect to the position of the BMU on the SOM as well as the model realizations that were associated to it.

4 Discussion and conclusions

The performance measures that are linked to the map in Fig. 2 already indicate that very individual properties of the training data time series can be attributed to each element of the Self-Organizing Map. Furthermore, from the patterns of the performance measures on Fig. 2 it can be seen that certain correlation structures inherent to these statistical measures

appear to be reflected by the map. Henceforth, we deduce that the information that can be extracted by these aggregating statistical measures is assimilated and preserved by the SOM. The findings with respect to Fig. 3, corroborated by Fig. 4 and Table 3, demonstrate that the SOM application is capable of revealing information about parameter sensitivities and, to a certain degree, parameter interactions. We consider these results an indication of the high discriminative power of the SOM application with respect to the characteristics of different simulated discharge time series. This is because we were not able to obtain similar findings with traditional methods that are based on the evaluation of performance measures, e.g. parameter response surfaces for different objective functions.

These useful aspects of the method are complemented by the findings of the second experiment: it demonstrates that the information which is processed by the SOM allows differentiating the spectrum of model realizations, given with the Monte-Carlo data, such that a rather narrowly confined set of model time series which are similar to the observed time series can be identified. Based on a visual examination of the resulting hydrographs, the SCE optimization seems to slightly outperform the results of the SOM. Standard

objective function values for these time series also suggest a higher accuracy of the results obtained by application of the SCE algorithm. In contrast to the SCE algorithm the resolution of the SOM method is dependent upon the number of model time series that participated in its training. It should be borne in mind that, compared to the number of model parameters, the results given in Fig. 7c are still based on a rather small number of model data items for the SOM training. The results could therefore improve when a more densely sampled dataset is used for the training. Nevertheless the model realizations that have been attributed to the BMU already exhibit qualities similar to the result which was based on optimization with the SCE algorithm. The differences between these realizations might further be attributed to the fact that the SOM training does not tend to put emphasis on particular hydrograph features, which however can be expected when using RMSE as optimization criterion. A strong point of the SOM procedure is its ability to provide a number of alternative model realizations that approximate the measured time series equally well. This allows for confining the parameter space to a region that closely represents a measured data set and renders the SOM potentially useful for future model identification applications. Most notably it has to be pointed out that the SOM method does not depend on aggregating statistical measures. Consequently the "similarity" represented in the SOM is not directly quantifiable in traditional terms. Instead, it rather accounts for the complexity that is inherent to time series data and which cannot be reduced to a rank number. Although the method is deterministic and the results are entirely reproducible, the resulting time series (Fig. 7c) can be further judged only subjectively. The fact that the experiments yielded identical results when the training was carried out using discharge data underpins the stability of the SOM method.

The discriminatory power of the SOM that has been demonstrated in this article also highlights that uncertainty induced by the properties of the performance measure should be included in the discussion of model uncertainties and equifinality, because any statement on model behaviour depends on our possibilities to differentiate between model time series.

Edited by: F. Laio

References

Alhoniemi, E., Hollmén, J., Simula, O., and Vesanto, J.: Process Monitoring and Modeling using the Self-Organizing Map, Integr. Comput.-Aid. E., 6, 3–14, 1999.

Allinson, N. M. and Yin, H.: Self-Organising Maps for Pattern Recognition, in: Kohonen Maps, edited by: Oja, E. and Kaski, S., Elsevier, Amsterdam, 111–120, 1999.

Ambroise, B., Perrin, J. L., and Reutenauer, D.: Multicriterion validation of a semidistributed conceptual model of the water cycle in the Fecht Catchment (Vosges Massif, France), Water Resour. Res., 31, 1467–1482, 1995.

Beven, K. J. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, 1992.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, Water Resour. Res., 36, 3663–3674, 2000.

Cho, J.: Multiple Modelling and Control of Nonlinear Systems with Self-Organizing Maps, PhD Thesis, University of Florida, 130 pp., 2004.

Clemen, T.: Zur Wavelet-gestützten Validierung von Simulationsmodellen in der Ökologie, Dissertation, Institut für Informatik und Praktische Mathematik, Christian-Albrechts-Universität, Kiel, 111 pp., 1999.

Duan, Q., Gupta, V. K., and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, J. Optimiz. Theory App., 76, 501–521, 1993.

Franks, S. W., Gineste, P., Beven, K. J., and Merot, P.: On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process, Water Resour. Res., 34, 787–797, 1998.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour. Res., 34, 751–764, 1998.

Gupta, H. V., Sorooshian, S., Hogue, T. S., and Boyle, D. P.: Advances in Automatic Calibration of Watershed Models, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application, AGU, Washington D.C., 9–28, 2003.

Haykin, S.: Neural networks – a comprehensive foundation, 2nd ed., New Jersey, 842 pp., 1999.

Hsu, K.-l., Gupta, H. V., Gao, X., Sorooshian, S., and Imam, B.: Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis, Water Resour. Res., 38, 1302, doi:10.1029/2001WR000795, 2002.

Hydrotec: Rainfall-Runoff-Model NASIM – program documentation (in German), Hydrotec Ltd., Aachen, 579 pp., 2005.

Kalteh, A. M. and Berndtsson, R.: Interpolating monthly precipitation by self-organizing map (SOM) and multilayer perceptron (MLP), Hydrolog. Sci. J., 52, 305–317, 2007.

Kohonen, T.: Self-Organizing Maps, 3rd ed., Information Sciences, Berlin, Heidelberg, New York, 501 pp., 2001.

Lamb, R., Beven, K., and Myrabø, S.: Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model, Adv. Water Resour., 22(4), 305–317, 1998.

Hydrol. Earth Syst. Sci., 12, 657–667, 2008

www.hydrol-earth-syst-sci.net/12/657/2008/

Lane, S. N.: Assessment of rainfall-runoff models based upon wavelet analysis, Hydrol. Process., 21, 586–607, 2007.

Legates, D. R. and McCabe Jr., G. J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35, 233–241, 1999.

Lischeid, G.: A decision support system for mountain basin management using sparse data, EGU General Assembly 2006, Vienna, Geophys. Res. Abstr., 8, EGU06-A-04223, 2006.

Montanari, A. and Toth, E.: Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, Water Resour. Res., 43, W05434, doi:10.1029/2006WR005184, 2007.

Peeters, L., Bação, F., Lobo, V., and Dassargues, A.: Exploratory data analysis and clustering of multivariate spatial hydrogeological data by means of GEO3DSOM, a variant of Kohonen's Self-Organizing Map, Hydrol. Earth Syst. Sci., 11, 1309–1321, 2007, http://www.hydrol-earth-syst-sci.net/11/1309/2007/.

Principe, J. C., Wang, L., and Motter, M. A.: Local dynamic modeling with self-organizing maps and applications to nonlinear system identification and control, P. IEEE, 86, 2240–2258, 1998.

Schütze, N., Schmitz, G. H., and Petersohn, U.: Self-organizing maps with multiple input-output option for modeling the Richards equation and its inverse solution, Water Resour. Res., 41, W03022, doi:10.1029/2004WR003630, 2005.

Seibert, J.: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, Hydrol. Earth Syst. Sci., 4, 215–224, 2000, http://www.hydrol-earth-syst-sci.net/4/215/2000/.

Simula, O., Vesanto, J., Alhoniemi, E., and Hollmén, J.: Analysis and Modeling of Complex Systems Using the Self-Organizing Map, in: Neuro-Fuzzy Techniques for Intelligent Information Systems, edited by: Kasabov, N., and Kozma, R., Physica Verlag (Springer Verlag), 3–22, 1999.

Vesanto, J.: Using the SOM and Local Models in Time-Series Prediction, Workshop on Self-Organizing Maps (WSOM'97), Espoo, Finland, 1997, 209–214, 1997.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J.: SOM Toolbox for Matlab 5, Helsinki University of Technology, Espoo, 60 pp., 2000.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resour. Res., 39(4), 1214, doi:10.1029/2002WR001746, 2003.

Wagener, T., Wheater, H. S., and Gupta, H. V.: Identification and Evaluation of Watershed Models, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application, AGU, Washington D.C., 29–47, 2003.

Wagener, T., Wheater, H. S., and Gupta, H. V.: Rainfall-Runoff Modelling in Gauged and Ungauged Basins, Imperial College Press, London, 306 pp., 2004.

Willmott, C. J.: On the validation of models, Phys. Geogr., 2, 184–194, 1981.

Willmott, C. J.: Some Comments on the Evaluation of Model Performance, B. Am. Meteorol. Soc., 63, 1309–1313, 1982.

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

Appendix B

Hydrology and Earth System Sciences

Mapping model behaviour using Self-Organizing Maps

M. Herbst[1]**, H. V. Gupta**[2]**, and M. C. Casper**[1]

[1]Department of Physical Geography, University of Trier, Germany
[2]Department of Hydrology & Water Resources, University of Arizona, Tucson, USA

Abstract. Hydrological model evaluation and identification essentially involves extracting and processing information from model time series. However, the type of information extracted by statistical measures has only very limited meaning because it does not relate to the hydrological context of the data. To overcome this inadequacy we exploit the diagnostic evaluation concept of Signature Indices, in which model performance is measured using theoretically relevant characteristics of system behaviour. In our study, a Self-Organizing Map (SOM) is used to process the Signatures extracted from Monte-Carlo simulations generated by the distributed conceptual watershed model NASIM. The SOM creates a hydrologically interpretable mapping of overall model behaviour, which immediately reveals deficits and trade-offs in the ability of the model to represent the different functional behaviours of the watershed. Further, it facilitates interpretation of the hydrological functions of the model parameters and provides preliminary information regarding their sensitivities. Most notably, we use this mapping to identify the set of model realizations (among the Monte-Carlo data) that most closely approximate the observed discharge time series in terms of the hydrologically relevant characteristics, and to confine the parameter space accordingly. Our results suggest that Signature Index based SOMs could potentially serve as tools for decision makers inasmuch as model realizations with specific Signature properties can be selected according to the purpose of the model application. Moreover, given that the approach helps to represent and analyze multi-dimensional distributions, it could be used to form the basis of an optimization framework that uses SOMs to characterize the model performance response surface. As such it provides a powerful and useful way to conduct model identification and model uncertainty analyses.

1 Introduction

Diagnostic model evaluation and identification aim at elucidating the extent to which the model is able to represent the observed system behaviour and identifying the reasons why its response is inconsistent with the observations (Gupta et al., 2008). This task – a fundamental step in the iterative model building and identification process – requires adequate tools that guide the modeller towards isolating the causes of "unsatisfactory" model behaviour such that changes to the parameters or the model structure can be made accordingly (Wagener et al., 2003b; Gupta et al., 2008). Model evaluation is commonly based on the comparison of model generated input-state-output simulations with observed historical data. In its simplest, but most powerful, form this is done visually in the course of manual calibration, because a trained expert is able to simultaneously discern various characteristics of the data and relate them to the hydrological context.

Already considerable progress has been made in formalizing the diagnostic process of model identification. Recognizing that manual calibration usually involves the evaluation of a number of different aspects of the time series, it follows that model calibration is inherently a multi-objective problem that requires the evaluation of more than one performance measure to ensure sufficient discriminatory power, even if the model produces only a single output time series (Gupta et al., 1998). And of course, complementary sources of information such as ground water measurements or soil moisture observations, among others, can be used in this process (Ambroise et al., 1995; Franks et al., 1998; Lamb et al., 1998; Seibert et al., 2000; Gallart et al., 2007). By measuring the model performance during different stages of the model response Boyle et al. (2000) and Wagener et al. (2003a) were able to relate model performance to individual model components. In this way, it became possible to extract information regarding the functioning and capabilities of the model from its output time series, which consequently allowed testing

Correspondence to: M. Herbst
(herbstm@uni-trier.de)

of the underlying model hypothesis. Hence, the inability of a model to reproduce the whole time series with a single parameter set can be indicative of structural model errors (Gupta et al., 1998; Wagener et al., 2003b; Lin and Beck, 2007). However, errors in the forcing data as well as in the parameterization are also frequent causes of temporal parameter variability.

Even so, evaluation methods that rely on a quantification of the "difference" between the simulation and the measurements commonly resort to regression based statistical measures of fit as the primary method of information extraction (Legates an McCabe Jr., 1999; Willmott et al., 1981; Nash and Sutcliffe, 1970). The pitfalls of using such measures are well known (e.g. see Hall, 2001; Lane, 2007; Schaefli and Gupta, 2007). However, when dealing with model evaluation in a diagnostic sense, the nature of the 'information' we extract from the input-output data by means of these measures deserves critical attention. Gupta et al. (2008) point out that different types of information can be extracted depending on the context in which the data is placed. By putting data in the context of common statistical measures of fit, we allow primarily for a correlative evaluation of the data; e.g., the correlation coefficient informs about the percentage of the observed variance that can be explained by the model simulation, but conveys little or no information that relates directly to the hydrological context of model/data, e.g. a bias in the water balance, or in the velocity of the rainfall-runoff response.

Because variance-based statistical measures fail to extract and relate the information contained in the model time series data to characteristics that are interpretable and meaningful in the context of the hydrological theory, they offer only minimal diagnostic support when used at the front end of various model evaluation frameworks. In response to this, Gupta et al. (2008) and Yilmaz et al. (2008) recently proposed the concept of a diagnostic evaluation approach rooted in information theory, in which a set of measures is used to characterize various theoretically relevant system functions and process behaviours. The term "Signature Indices" is introduced to distinguish these measures from conventional variance based performance measures that lack relationships to the underlying hydrological theory. Yilmaz et al. (2008) propose a set of such Signature Indices to help guide the diagnostic process of model evaluation in a meaningful and interpretable way, inasmuch as these measures correspond to four major hydrological functions of the watershed: overall water balance, vertical redistribution, temporal redistribution and spatial redistribution.

In complementary work, Herbst and Casper (2008) propose a model evaluation approach that circumvents the low discriminatory power of statistical performance measures (Gupta et al., 2003), while maximally exploiting the potential information in the data, by use of the power of Self-Organizing Maps (SOMs) (Kohonen, 2001). A Self-Organizing Map consists of an unsupervised learning neu-

ral network algorithm that performs a non-linear mapping of the dominant structures present in a high-dimensional data field onto a lower-dimensional grid. The SOM has found diverse applications in fields such as pattern recognition, image analysis (Kohonen, 2001), exploratory data analysis (Kaski, 1997; Vesanto, 2000a) in geo-spatial (Lourenço, 2005) as well as hydrochemical data (Boogaard et al., 1998; Lischeid, 2006; Peeters et al., 2007), process monitoring (Alhoniemi et al., 1999; Simula et al., 1999), local time series modelling (Vesanto, 1997) and time series forecasting (Simon et al., 2005). Applications related to hydrological modelling remain far less numerous, albeit rather diverse: Kalteh and Berndtsson (2007) use SOMs for the interpolation of monthly precipitation; Lin and Chen (2006) apply SOMs for regional precipitation frequency analysis; Schütze et al. (2005) use an extension of the SOM to approximate the Richards equation and its inverse; Huang et al. (2003), and similarly Rajanayaka et al. (2003), apply SOMs to cluster and categorize soil data sets in a framework for estimating model parameters in data-sparse areas; Chang (2001) explores the use of SOMs to infer physical and hydraulic soil properties based on remotely sensed brightness temperature data; Mele and Crowley (2008) apply an SOM to examine the interrelationship between different bio-indicators and hydrological soil properties.

In the context of watershed modelling, Hsu et al. (2002) successfully performed daily streamflow predictions using an SOM to identify different rainfall-runoff stages to which local regression functions are assigned – a similar approach is used by Abramowitz et al. (2006) to characterize the systematic component of land surface model output error in a flux correction technique for land surface models (Abramowitz et al., 2007). Abramowitz and Gupta (2008) also use an SOM to evaluate multi-model independence. A further application to land surface modelling is presented by Abramowitz et al. (2008) who use SOM to cluster time steps with similar meteorological condition in order to examine the conditional bias of land surface models. Reusser et al. (2008) present another interesting approach in the field of model evaluation. They analyze the temporal dynamics as well as the type of model error by using SOM for the assessment of multiple performance measures from a moving window approach.

In a rather general sense, the SOM is one of several methods that can be used for time series clustering (Warren Liao, 2005). Recently, Herbst and Casper (2008) used an SOM-based approach in this sense, to obtain a topologically ordered classification and clustering of the temporal patterns present in model outputs obtained from Monte-Carlo simulations. This allowed the authors to differentiate the spectrum of simulated time series with a high degree of discriminatory power. The work shows that the SOM can provide insights into parameter sensitivities, while helping to constrain the model parameter space to the region that best represents the measured time series. However, the Herbst and Casper (2008) approach shares a shortcoming of methods based on

Table 1. NASIM model parameters of the Monte-Carlo simulation with their respective parameter bounds.

Name	Description	Unit	Lower Bound	Upper Bound
RetBasis	Storage coefficient factor for baseflow component	[–]	0.5	3.5
RetInf	Storage coefficient factor for interflow component	[–]	2.0	6.0
RetOf	Storage coefficient factor for surface runoff from unsealed surfaces	[–]	2.0	6.0
StFFRet	Storage coefficient factor for surface runoff from urban areas	[–]	2.0	6.0
hL	Horizontal hydraulic conductivity factor	[–]	2.0	8.0
maxInf	Maximum infiltration factor	[–]	0.025	1.025
vL	Vertical hydraulic conductivity factor	[–]	0.005	0.105

statistical goodness-of-fit measures – because the clustering into SOM nodes is not based on patterns/properties that have direct hydrological relevance, the map is not easy to interpret. Further, because the SOM training involves use of entire (potentially very long and/or high-dimensional) time series the computational costs are rather high.

The main goal of this paper is to explore how the Herbst and Casper (2008) SOM-based approach can be improved by linking it to the Signature Index concept (Gupta et al., 2008), which defines the similarity between data items in a more meaningful (hydrologically relevant) way. So, instead of working directly with the data time series, the SOM training is conducted on the set of Signature Indices introduced by Yilmaz et al. (2008), computed from each of the Monte-Carlo simulation time-series generated by the distributed conceptual hydrological model. In this context, it is interesting to note that the SOM based method presented in this paper in many aspects aims at similar objectives as the Parameter Identification Method based on the Localization of Information (PIMLI) developed by Vrugt et al. (2001, 2002). The PIMLI method attempts to find disjunctive subsets of the observations that contain the most information for the different model parameters. It thus provides a diagnostic instrument that helps differentiating between parameters and model concepts. In contrast to the SOM based method PIMLI is a sequential optimization methodology that approaches the model evaluation problem more from the observations than the model itself.

The manuscript is organized as follows; following a brief overview of the model and the data used (Sect. 2.1), we discuss how the hydrologically relevant Signature Indices are computed from model output time series (Sect. 2.2). Section 2.3 presents a brief discussion of the concept and properties of the Self-Organizing Map, followed by a discussion of the training process (Sect. 2.4). The results are presented in Sect. 3 followed by a discussion (Sect. 4) of possible reasons for the differences between the findings presented in this contribution and the results previously published by Herbst and Casper (2008). We conclude with suggestions for other possible applications, and try to illuminate the method in the context of existing model evaluation and optimization methods.

2 Methods

2.1 Model and data

This work employs the same model and data used by Herbst and Casper (2008). Monte-Carlo simulations of hourly stream flow, over a period of approximately two years, were generated using the distributed conceptual watershed model NASIM; see Hydrotec (2005) for details. The model uses a spatial discretization based on sub-catchments. For the "Schwarze Pockau" watershed preprocessing of spatial data resulted in 71 sub-catchments with a mean size of approximately $1.8 \, \text{km}^2$. These are further subdivided into spatially homogeneous units with respect to soil and land use. Each of these elementary spatial units is again vertically divided into soil layers. All lateral flow components that result from the processes within the elementary units are aggregated on the sub-catchment scale. The NASIM parameters examined in this study (Table 1) are unit less factors that modify internal parameter values that are either based on global default values or have been determined individually for each sub-basin in the course of the spatial data preprocessing: The internal values modified by RetOf are determined in the course of the preprocessing depending on the slope in each sub-basin, while the internal RetInf, RetBas, StFFRet are set to global values. The internal values of maxInf as well as vL are determined according to soil type. The model has been distributed commercially since the mid-eighties, and has found widespread application in water resources management throughout Germany.

In this work, we adopt the decision-maker's point of view, treat the model as a black-box, and try to understand its functioning via the methods presented in this paper. The test watershed is the low-mountain range $129 \, \text{km}^2$ catchment "Schwarze Pockau" Saxony (Germany), a tributary of the Freiberger Mulde (Elbe sub-basin) situated near the border to Czech Republic.

The data consist of hourly precipitation and streamflow measurements at gaging station "Zöblitz". We focus on the period from 1 November 1994 to 28 October 1996 and use the preceding one-year as a warm-up period, for a total of 17 472 simulation time steps. A total of 4000 simulation runs

were generated by randomly varying seven of the model parameters (Table 1), all related to the soil water balance and vertical redistribution of flow components, over their feasible ranges. Because appropriate prior information on parameter (or rather factor) distributions was missing uniform distributions were assumed. The variation of these factors during the Monte-Carlo simulation was performed with global values for all sub-catchments. The ranges for the free parameter factors and the values for the fixed parameter factors were set based on prior knowledge acquired via manual expert calibration to the test watershed. We assume that these values represent the plausible parameter space for this watershed with very high probability.

2.2　Signature indices

The Signature Indices used in this work were designed with a view to providing generally applicable and meaningful measures of model performance that can provide information helpful in detecting and isolating causes of model inadequacies, thereby providing guidance towards model improvements (Yilmaz et al., 2008). Unlike standard statistical measures of fit these Signature Indices are rooted in the context of the hydrological theory underlying our conceptual representation of the watershed. The reason for multiple indices is that each is designed to target different (complementary) aspects of model/system behavior (see Gupta et al., 2008). Following the concepts and mathematical formulation presented by Yilmaz et al. (2008), we focus on five Signature Indices that provide information about the following intrinsic watershed characteristics:

– overall water balance

– vertical soil moisture redistribution

– behaviour of long-term baseflow

Because none of the model parameters investigated in this paper control the streamflow routing and timing, we do not consider here the fourth intrinsic water shed characteristic – temporal redistribution of flow – but details can be found in Yilmaz et al. (2008).

The first Signature Index focuses on the long-term input-output behaviour of the system (volume balance), and therefore measures the percent bias in overall runoff (%BiasRR). This index is strongly controlled by the evapotranspiration process and any factors that influence the amount of water available for evapotranspiration. The index should therefore be sensitive to the model components and parameters that control these processes. It is computed as

$$\%\mathrm{BiasRR} = \frac{\sum_t (\mathrm{Qsim}_t - \mathrm{Qobs}_t)}{\sum_t \mathrm{Qobs}_t} \times 100. \tag{1}$$

where Qsim and Qobs denote the simulated and the observed flow respectively. The next four Signature Indices are derived

from the flow exceedance probability curve (also known as flow duration curve, FDC), to represent watershed characteristics that act on short to intermediate time scales. Because the FDC characterizes a watershed's tendency to produce flows of different magnitudes, it is informative regarding the vertical redistribution of water within the soil column. Different sections of the flow duration curve can therefore be associated with the occurrence of fast, intermediate and slow runoff responses. Because construction of the FDC involves loss of timing information, these indices should be generally insensitive to output timing errors. For catchments with quick (slow) runoff response we expect the FDC mid-segment slope to be steeper (flatter).

The percent error in the FDC mid-segment slope is given by %BiasFDCm:

$$\%\mathrm{BiasFDCm} = \tag{2}$$
$$\frac{\left(\log(\mathrm{Qsim}_i) - \log(\mathrm{Qsim}_j)\right) - \left(\log(\mathrm{Qobs}_i) - \log(\mathrm{Qobs}_j)\right)}{\left(\log(\mathrm{Qobs}_i) - \log(\mathrm{Qobs}_j)\right)} \times 100$$

where i and j denote the lower and the upper threshold of flow exceedance probability that define the mid-segment of the flow duration curve. The volume of water corresponding to the high flow segment is also useful in a diagnostic context; this property is measured as the percent error in FDC high-segment volume %BiasFHV:

$$\%\mathrm{BiasFHV} = \frac{\sum_h (\mathrm{Qsim}_h - \mathrm{Qobs}_h)}{\sum_h (\mathrm{Qobs}_h)} \times 100 \tag{3}$$

where h denotes the indices of all discharge values with exceedance probabilities lower than 0.02. Similarly, the volume of water corresponding to long-term base flow is measured as the percent error in the FDC low-segment volume %BiasFLV:

$$\%\mathrm{BiasFLV} = \tag{4}$$
$$\frac{\sum_{l=1}^{L} (\log(\mathrm{Qsim}_l) - \log(\mathrm{Qsim}_L)) - \sum_{l=1}^{L} (\log(\mathrm{Qobs}_l) - \log(\mathrm{Qobs}_L))}{\sum_{l=1}^{L} (\log(\mathrm{Qobs}_l) - \log(\mathrm{Qobs}_l))} \times 100$$

where l denotes the index of the discharge values within the boundaries of the low flow segment with the index of its lowest value being L. Note that this index is computed according to Yilmaz et al. (2008) using a log transform of the flows to increase the sensitivity to very low flows. In principle, this index should be sensitive to components/parameters that influence base flow recession rates, as well as those that control the demand for evapotranspiration loss from the stream and/or the lower zone storage. The fourth FDC-based Signature Measure is the percent bias in median of the log transformed discharges %BiasFMM

$$\%\text{BiasFMM} = \tag{5}$$

$$\frac{\log(\text{Qsim}_{\text{median}}) - \log(\text{Qobs}_{\text{median}})}{\log(\text{Qobs}_{\text{median}})} \times 100$$

where $\text{Qsim}_{\text{median}}$ and $\text{Qobs}_{\text{median}}$ denote the median value of the observed and the simulated flow respectively. %BiasFMM characterizes differences in the mid-range flow levels between simulated and observed discharge. We observed, during this study, that this index is sensitive to errors in the recession limb of the flow hydrograph, specifically in the region of its inflection point.

Note that these indices achieve diagnostic value because a) the functions they measure must be represented in the model to enable proper reproduction of the system behaviour and b) they relate to aspects of the model structure/behavior that act on different time scales. However, we make no claim that this set of indices is comprehensive in representing the various characteristics that describe the behaviour of a watershed; in this regard, much research on Signature Index design and selection still needs to be done.

Based on the examination of the FDC for the observed discharge, the flow exceedance probability thresholds used in the definition of %BiasFHV, %BiasFDCm and %BiasFLV were subjectively set to i=20% and j=55% respectively. Accordingly, the index l in Eq. (4) denotes all discharge values with flow exceedance probabilities higher than 55%. Based on Eq. (1)–(5) the characteristics of a simulated time series of discharges can be described in a hydrologically meaningful way by a vector consisting of five Signature Indices.

2.3 The Self-Organizing Map

The Self Organizing Map is an artificial neural network technique designed to help in extracting structure from high-dimensional data sets (e.g. see Kohonen, 2001; Haykin, 1999; Kaski, 1997). Kohonen (2001) provides an exhaustive discussion of the algorithm and its properties. An SOM consists of a regular grid G of k neurons; in our case the neurons were arranged on a two-dimensional hexagonal grid, although other variants are common (Kohonen, 2001). Let X represent a set of Signature Index vectors, $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T \in \Re^n$ with n=5, that has been calculated according to Sect. 2.2 for a set of simulated discharge time series. Correspondingly, each neuron $i = 1\ldots k$ of G is represented through a reference vector

$$\boldsymbol{m}_i = [\mu_{i1}, \mu_{i2}, \ldots, \mu_{in}]^T \in \Re^n \tag{6}$$

whose dimension n equals the number of elements in an input data vector $\boldsymbol{x} \in X$. Typically, the reference vectors $\boldsymbol{m}_i$ are initialized to small random values. However, to assure faster and more reliable convergence of the map, we initialize the $\boldsymbol{m}_i$ along the two greatest principal component eigenvectors of the data (Kohonen, 2001). The SOM is trained iteratively:

In the first step an input data item $\boldsymbol{x} \in X$ is randomly selected and the Euclidean distance

$$d_i = \sqrt{\sum_{j=1}^{n} (x_j - m_{ij})^2} \qquad i = 1\ldots k; j = 1\ldots n \tag{7}$$

between $\boldsymbol{x}$ and each reference vector $\boldsymbol{m}_i$ is computed (of course, any appropriate metric can be used as a measure of similarity). The "winning neuron" (also called the best-matching unit, BMU, of $\boldsymbol{x}$) is the map element c whose reference vector $\boldsymbol{m}_c$ has the smallest distance d_c to $\boldsymbol{x}$ and thus satisfies the condition

$$d_c = \min_i \{|\boldsymbol{x} - \boldsymbol{m}_i|\} \qquad i = 1\ldots k. \tag{8}$$

In the next step the reference vector $\boldsymbol{m}_c$ (which corresponds to our current best-matching unit) and all reference vectors of its neighbouring neurons are updated according to

$$\boldsymbol{m}_i(t+1) = \boldsymbol{m}_i(t) + \alpha(t)h_{ci}(t)[\boldsymbol{x}(t) - \boldsymbol{m}_i(t)] \tag{9}$$

where $\boldsymbol{m}_i(t)$ is the current weight vector at iteration step t. Thus, the rate of change for each node of the map is scaled by three factors: a) the difference $(\boldsymbol{x}(t) - \boldsymbol{m}_i(t))$ between the input data set $\boldsymbol{x}$ and the prototype vector $\boldsymbol{m}_i$ b) the size of a neighbourhood function h_{ci} which decreases monotonically to zero with t and with distance from the winning neuron and c) a learning rate factor $\alpha(t)$ which gradually lowers the height of the neighbourhood function as the iteration advances. For h_{ci} it is common to use the Gaussian function

$$h_{ci}(t) = \exp\left(-\frac{\|\boldsymbol{r}_c - \boldsymbol{r}_i\|^2}{2\sigma^2(t)}\right) \tag{10}$$

where $\sigma(t)$ defines the width of the topological neighbourhood, and both $\sigma(t)$ in Eq. (10) and $\alpha(t)$ in Eq. (9) decrease monotonically with t. Note that an exact choice of the function $\alpha(t)$ is not required (Kohonen, 2001). In this way, the SOM combines elements of competitive and adaptive learning. Repeated cycling through the training steps causes different nodes and regions of the map to be "tuned" to specific domains of the input space. Importantly, the enforced local interaction between the SOM nodes, implemented by Eq. (9), results in the map gradually developing an ordered and smooth representation of the input data space (Kaski, 1997).

As an alternative to the sequential approach expressed by Eq. (9), this work used Kohonen's "batch-training" algorithm (Vesanto, 2000b) to speed up the training process. Here, in each training step the data set is partitioned according to the Voronoi regions of the $\boldsymbol{m}_i$. Instead of sequentially running through all data items in each training cycle the whole data set X is presented to the map as a whole at

www.hydrol-earth-syst-sci.net/13/395/2009/

Hydrol. Earth Syst. Sci., 13, 395–409, 2009

each training cycle. The reference vectors are updated according to the weighted average of the data samples

$$\boldsymbol{m}_i(t+1) = \frac{\sum\limits_{l=1}^{N} h_{ci}(t)\boldsymbol{x}_l}{\sum\limits_{l=1}^{N} h_{ci}(t)} \tag{11}$$

where c is the index number of the BMU of data set $\boldsymbol{x}_l$, and N is the number of data samples. This variant of the training does not make use of the learning rate factor $\alpha(t)$.

The adaptive and competitive learning process of the SOM along with its ability to provide a visualization of the data mapped onto a two-dimensional grid confers properties to the SOM which make it especially interesting for exploratory data analysis (Kaski, 1997). It results in an ordered representation of complex (i.e. multi-dimensional) data in which items associated with neighbouring nodes of the grid are characterized by similar patterns/properties, which facilitates the perception of structures inherent to the data. Further, patterns that occur more frequently in the input space are mapped onto a larger area. Additionally, the final reference vectors form a discrete approximation of the input data distribution by representing local conditional expected values of the data. They can thus be seen as essentially equivalent to a discretized form of principal curves (Kaski, 1997; Haykin, 1999). Consequently, depending on the scope of its application, SOM can be used to disclose the clustering structure of the data, or be interpreted as a dimensionality reduction or data compression method.

Finally, the SOM also allows the projection of an input data item $\boldsymbol{y}$ (in our study a vector of Signature Indices) that has not been part of the training data onto the output space. This means that according to Eq. (8) the neuron $c(\boldsymbol{y})$ with reference vector $\boldsymbol{m}_{c(y)}$ is determined that satisfies the condition

$$\left\|\boldsymbol{y} - \boldsymbol{m}_{c(y)}\right\| = \min_i \left\{\left\|\boldsymbol{y} - \boldsymbol{m}_i\right\|\right\} i = 1\ldots k. \tag{12}$$

Neuron $c(\boldsymbol{y})$ then represents the domain of input data patterns (i.e. Signature Index patterns) from X that is most similar to $\boldsymbol{y}$. In turn, the data items $X_c \subset X$ which are attributed to c are those among the training data items that are most similar to $\boldsymbol{y}$ with respect to the criterion given by Eq. (7) and Eq. (12).

2.4 Data preparation and training

In this work, the training data set X for the SOM was obtained by computing the five aforementioned Signature Indices for each of the 4000 model output time series obtained with NASIM as described in Sect. 2.1 and 2.2. Prior to the training, each Signature Index was normalized to a value having zero mean and variance of one using the linear transformation

$$x' = (x - \bar{x})/\sigma_x \tag{13}$$

so that high index values do not exert a disproportionate influence on the training via Eq. (7). The number of neurons, i.e. map units, was determined using the heuristic equation $m = 5\sqrt{N}$ where N=4000 is the total number of data items in X. The side lengths of the map are calculated based on the ratio of the two biggest eigenvalues of the covariance matrix of X (Vesanto et al., 2000). The training was conducted in two consecutive steps. First, a coarse training period of 500 iterations was performed, using a large radius for the neighbourhood function. Subsequently a fine tuning period comprising 10000 training cycles was performed using a small neighbourhood radius. A hexagonal grid was chosen to help preserve proper topologic relationships. A simple measure of the "quality" of the map is the "quantization error" (Vesanto, 2000a), which represents the average distance of each data vector $\boldsymbol{x}_p$ to its corresponding BMU $\boldsymbol{m}_{c(p)}$, calculated as:

$$\bar{d} = \frac{1}{N} \sum_{p=1}^{N} \left\| \boldsymbol{x}_p - \boldsymbol{m}_{c(p)} \right\| \tag{14}$$

where N is the total number of data items X used for training, and p is the index of the data items.

According to Eqs. (1)–(5) the Signature Index values associated with the time series of observed discharges Qobs maps as $\boldsymbol{y}=[0\,0\,0\,0\,0]^T$ into the Signature Index space. Using the transformation Eq. (13) that has been applied to the training data and Eq. (12) the map element $c(\boldsymbol{y})$ and its related data items X_c were determined that correspond to the Signature Index vector of the time series of observed discharges $\boldsymbol{y}=[0\,0\,0\,0\,0]^T$. We will refer to $c(\boldsymbol{y})$ as the "best-matching unit" (BMU) of the observed data.

As an independent reference point for this result, we used the Shuffled Complex Evolution optimization algorithm (Duan et al., 1992) to find a model realization (parameter set) that minimizes the Root Mean Squared Error (RMSE) of the simulated discharge time series. The SCE-UA algorithm was run with 5 complexes (5 points each) and a maximum of 10 000 iterations. For successful termination a change of less than 0.05 percent of the RMSE in three consecutive loops was imposed.

As a further reference point we conducted a SOM training according to Herbst and Casper (2008), using the discharge time series from the Monte-Carlo simulation as training data. The preprocessing was carried out similarly, applying the abovementioned transformation Eq. (13) to each time step of the training data set. For further details regarding this training please see Herbst and Casper (2008).

3 Results

3.1 Properties of the map

The resulting map of Signature Indices consists of $25 \times 13 = 325$ neurons, a number considerably smaller than the
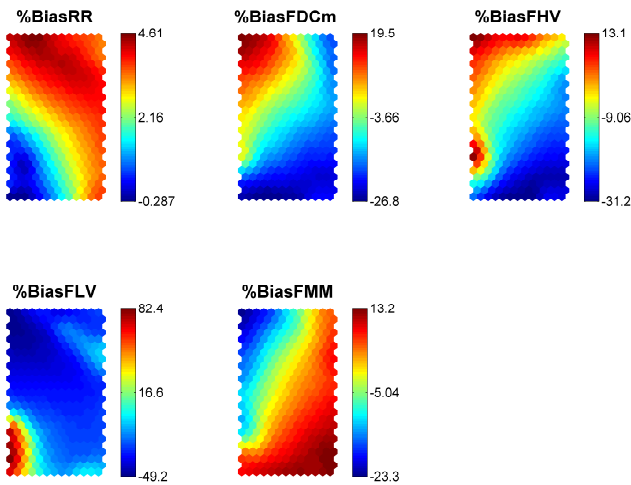
Fig. 1. Distribution of reference vector Signature properties. Note that in each plot the colours are scaled individually.
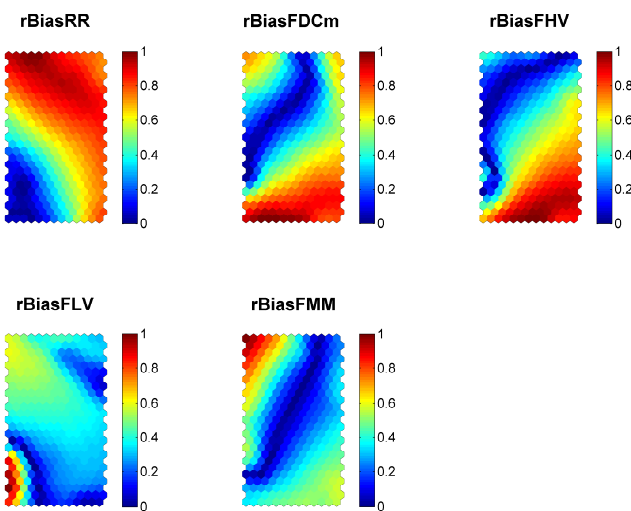

Fig. 2. Component planes of the SOM rescaled according to the normalized absolute values of the reference vectors clearly highlight the locations of individual Signature Index optima on the map.

total number of data items used for training. Consequently, every neuron represents a set of simulation runs and their respective Signature Index patterns. As the reference vectors of the map are readily accessible, the properties of the individual nodes, with regard to their Signature Index characteristics, can be evaluated: The de-normalized reference vectors represent the "prototype patterns" by which the individual neurons are activated. Further, we take advantage of the fact that the input data items attributed to each neuron via the training are referenced. Because each input data item is linked to a parameter set and its resulting simulated time series, the neurons of the map can be evaluated with respect to the model parameters and the properties of the time series.
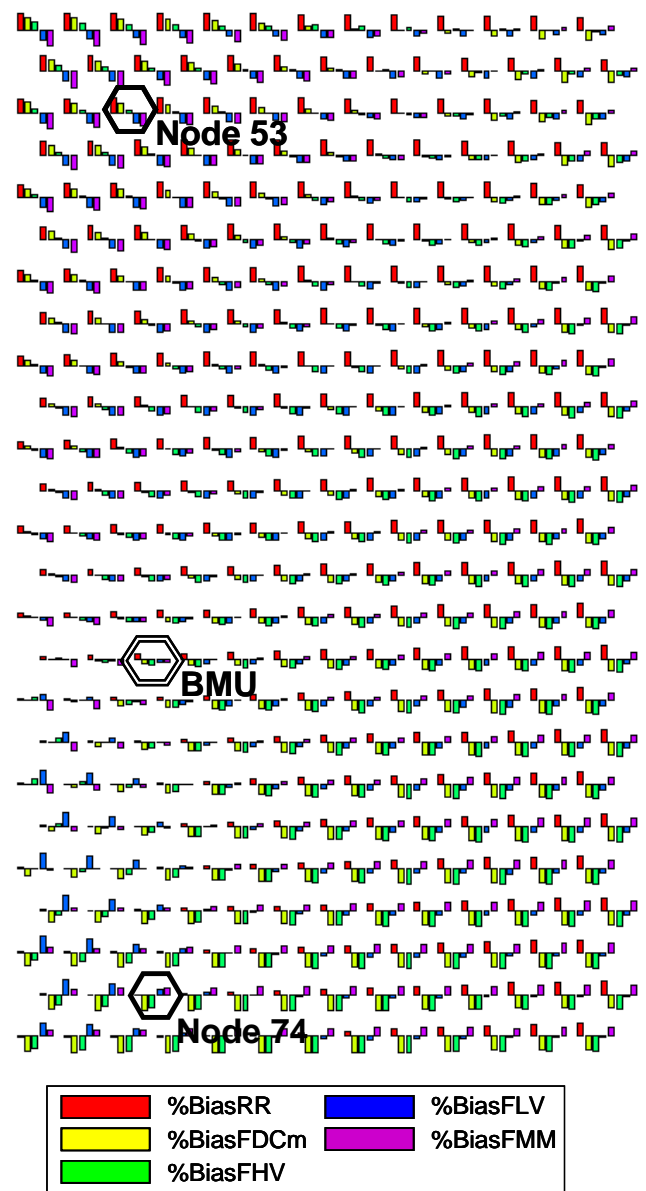
Fig. 3. The distribution of Signature Index properties on the map displayed as bar-plots and the location of the best-matching unit (BMU) on the map. The bars are individually scaled over the range of each Signature Index such that the topmost position of each bar marks its absolute maximum and vice versa. Details on properties of the model realizations projected onto node 53 and 74 are exemplified in Fig. 4.

In Fig. 1 the Signature properties of the reference vectors are visualized using component planes, i.e. by means of congruent, colour coded subplots that display the distribution of each reference vector component (i.e. each Signature Index) separately. Note that in each plot the colours are scaled individually. These distributions represent the Signature Index pattern of the Monte-Carlo model realizations. Figure 1 also highlights the trade-offs between different Signature Indices
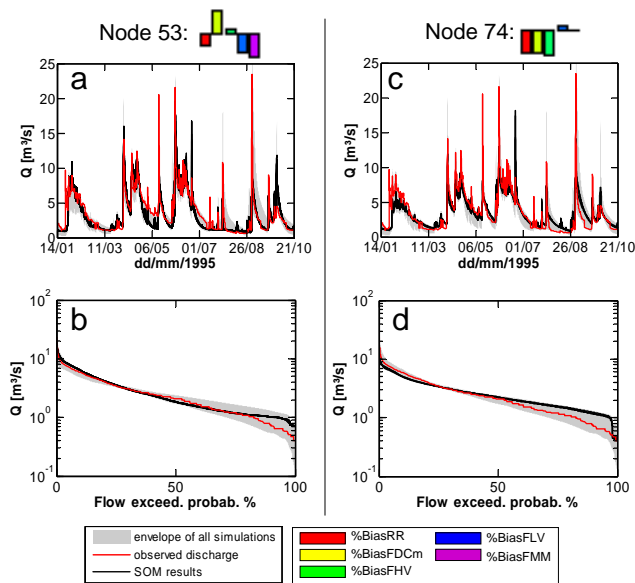
www.hydrol-earth-syst-sci.net/13/395/2009/

Hydrol. Earth Syst. Sci., 13, 395–409, 2009

Fig. 4. Examples for Signature Index properties on different locations of the map (see Fig. 3): Comparison of hydrographs and flow duration curves for the observed discharge and the model realizations projected onto each of the two nodes.

Fig. 5. Mean values of each model parameter for the simulations projected onto the individual map elements.

and therefore reveals which capabilities in reproducing the characteristics of the measured time series are mutually exclusive. To accentuate the location of individual Signature Index optima on the map Fig. 2 shows the component planes rescaled according to normalized absolute values of the reference vectors. Here, it becomes evident that the model optima, with respect to the Signature Indices, are in part mutually exclusive. Figure 3 further summarizes the characteristics of the reference vectors: here the Signature Indices of each reference vector are presented as bar-plots where each of the five color-coded bars corresponds to one of the Signature Indices. Note that the bars are individually scaled over the pertinent range of each Signature Index such that the topmost position of each bar marks its absolute maximum and vice versa. From Fig. 3 it can easily be discerned that the model realizations for which all Index values are closest to zero are projected to the neurons located around the left lower third of the map. (The location of the BMU according to Sect. 2.4 will be determined in the following section). The lower and lower right region of the map is dominated by model realizations with negative Index values regarding %BiasFDC and %BiasFHV which gradually switch over to positive values towards the upper left segment of the map. It is evident from Fig. 1 and 3 that the SOM provides a smooth mapping of the Signature Index properties such that neighbouring locations on the map bear similar characteristics with respect to the Signature Indices. The simulations that are represented by the neurons located in the upper left corner of the map (e.g. node 53 in Fig. 3) show positive values of %BiasRR, %BiasFDCm and %BiasFHV while at the same time
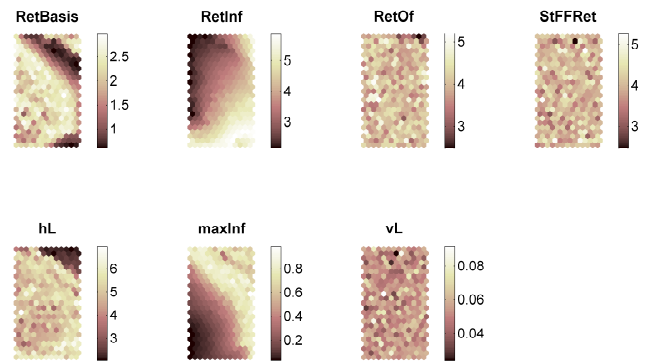
the values of %BiasFLV and %BiasFMM are among the lowest of the training data set. The hydrograph- and exceedance probability plots in Fig. 4a and b confirm that such Signature characteristics are indicative of model realizations with runoff components that react predominantly faster than the measured discharge time series whereas the quick recession of flow translates into volume errors in the intermediate flow components (i.e. higher and narrower flow hydrographs following storm events). Additionally, the positive deviation in overall water balance compared to the measured streamflow (i.e. high %BiasRR) combined with underestimation of low-flow volume corroborates that, on average, too much volume of flow is allocated to peak discharges. Note that the barplots as well as the component planes (Fig. 1) represent the relative scale of the Signature Indices.

Model realizations projected onto a neuron in the lower left part of the map (e.g. node 74 in Fig. 3), on the other hand, display a very different behaviour, as exemplified in Fig. 4c and d: From the combination of negative %BiasFDCm and %BiasFHV in conjunction with positive %BiasFLV and close to zero %BiasRR, it immediately becomes comprehensible that these simulations are marked by overall underestimation of peak flows, delayed recession of flow and overestimation of volume in the low flow component.

We further calculated the mean values of each model parameter of the simulations that have been attributed to the individual map elements: Figure 5 shows the distribution of the model parameter values that correspond to the Signature Indices given in Fig. 1 and 2. Here, conformities or similarities with the patterns of Fig. 1 will point at high correlations between parameter values and Signature Indices. Because the Signature Indices represent meaningful hydrological characteristics, these figures indirectly reveal the function of each parameter in the hydrological context. (Notice that in each lattice the positions of the neurons remain identical). Comparing the patterns of Signature Indices and parameter values over the map suggests that the parameter maxInf (= "max. infiltration rate") predominantly exerts a high influence on the
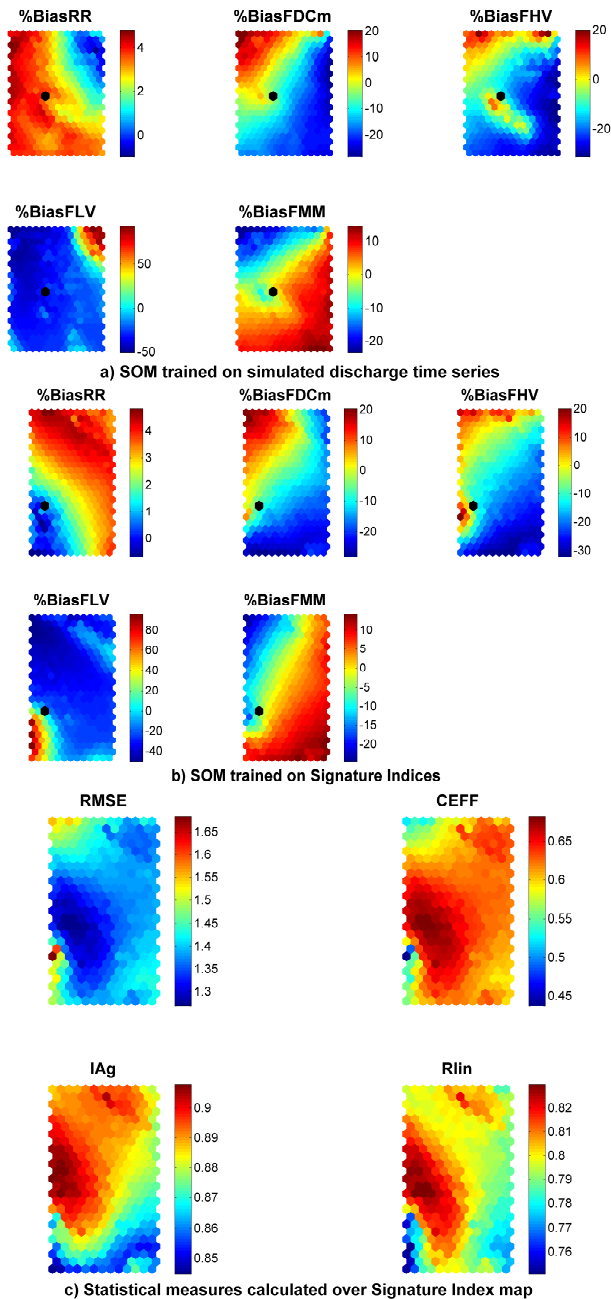
a) SOM trained on simulated discharge time series

b) SOM trained on Signature Indices

c) Statistical measures calculated over Signature Index map

Fig. 6. Comparison of SOM trained on simulated discharge time series **(a)** and the SOM trained on Signature Indices **(b)** by means of Signature Index mean values for the simulations which are attributed to each node. The black dot indicates the position of the BMU of the measured discharge time series. In **(c)** the SOM based on Signature Indices is represented by the mean values of four statistical performance measures – the root mean square error (RMSE), the Nash-Sutcliffe coefficient of efficiency (CEFF), Willmott's Index of Aggreement (IAg) and the correlation coefficient (Rlin) – that were calculated calculated over the sets of model realizations that are projected onto each neuron.

runoff bias %BiasRR. Likewise, parameter RetInf obviously controls the velocity of the runoff reaction, i.e. the vertical distribution of flow components, and the volumina allocated to high flows as can be observed by the good correspondence with the pattern of %BiasFDCm and %BiasFHV. The parameters RetOf, StFFRet and vL on the other hand display an irregular, "noisy" pattern in the distribution of parameter values over the map which does not correspond to any of the patterns present in the Signature Indices. Parameter hL shows the same irregular pattern over wide parts of the map but also presents isolated areas where the parameter values seem to be closely related to the ordered structure of the map (whose structure is governed by Signature Index properties). The aforementioned results lead us to infer that this pattern is likely to be indicative of partial parameter sensitivity, i.e. parameter interaction involving threshold behaviour. The pattern displayed by RetOf, StFFRet and vL, on the other hand, can only be explained if changes made to these parameters either did not relate to any of the Signature Indices or if the effect of these parameters were strongly tied to other parameter values in a highly nonlinear way.

Finally, we compare the ordering principle of the time series map computed by Herbst and Casper (2008) to the current map based on Signature Indices. For the former case, the mean values of the individual Signature Indices are calculated over the sets of time series that are projected onto each neuron of the map. Similar to the component plane visualization in Fig. 1 these mean Signature values are subsequently used to colour the map (Fig. 6a). Accordingly, the mean values of the Signature Indices are determined for the neurons of the Signature Index map in order to allow for proper comparability (Fig. 6b). (Note that the de-normalized reference vectors only refer to the properties of the neuron which are acquired during the training, whereas the element-wise mean values of the data items refer to the properties of the data which is projected onto these neurons). In Fig. 6 it is clearly noticeable that the arrangement of the data items on both maps agrees very well with respect to %BiasFDC, %BiasFMM and, to some extent, %BiasFHV. The components %BiasRR and %BiasFLV show almost identical patterns on both maps, which however are inversely arranged in each case.

In order to gain insight into the relationship between the ordering structure of the current SOM of Signature Indices and common statistical performance measures the mean values of the root mean square error (RMSE), the Nash-Sutcliffe coefficient of efficiency (CEFF; Nash and Sutcliffe, 1970), Willmott's Index of Aggreement (IAg; Willmott, 1981) and the correlation coefficient (R^2) are calculated over the sets of model realizations that are projected onto each neuron of the Signature Indices SOM. Not surprisingly, the results (Fig. 6c) reveal that the structure of the SOM based on Signature Indices translates into an ordered pattern of the four statistical objective functions on the map. These patterns, however, show close correlations. According to Fig. 2, 6b and c, it is
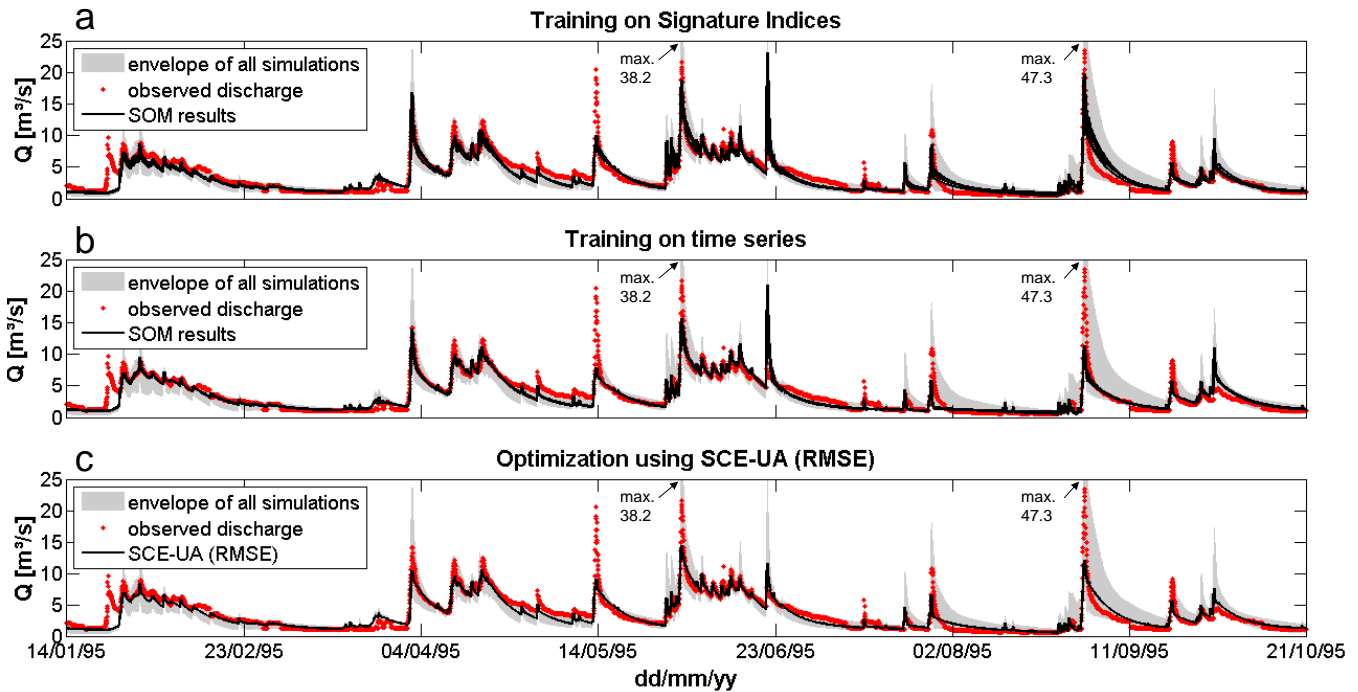
www.hydrol-earth-syst-sci.net/13/395/2009/

Hydrol. Earth Syst. Sci., 13, 395–409, 2009

Fig. 7. Comparison of model realizations for the BMU corresponding to the observed discharge time series. **(a)** The results of the SOM training on Signature Indices are contrasted to **(b)** the training on entire time series vectors and **(c)** the solution obtained by minimizing the RMSE using SCE-UA.

not possible to establish a clear relationship between the Signature Indices and the performance measures. With regard to the position of the optima on both mappings, it can be seen that the BMU in Fig. 6b does not coincide with the optima of the performance measures in Fig. 6c.

3.2 Projection of measured discharge time series onto the map

In Fig. 3 we also mark the location of the BMU of Qobs which was determined according to Sect. 2.4. Figure 3 shows that the BMU identifies a neuron with a reference vector that represents a compromise in minimizing all Signature Indices. In the following, we will further examine the model-realizations which are attributed to the BMU by virtue of the training process: Fig. 7a shows the 9 hydrographs retrieved from the BMU of the SOM training on Signature Indices along with the observed time series and the envelope of the entire Monte-Carlo data set. As can be seen, these hydrographs comprise rather similar model realizations, compared to the total range of Monte-Carlo simulations. In contrast, Fig. 7b shows the 7 time series obtained from identifying the BMU on a SOM that was trained on the entire time series of the model outputs (as per Herbst and Casper, 2008), without previously converting them to Signature Indices. Finally, Fig. 7c offers a comparison with the result obtained from minimizing the RMSE using SCE-UA (Duan et al., 1992). The time series that result from the training

on Signature Indices are not as densely bundled within the total range of Monte-Carlo simulations than their counterparts from Fig. 7b. However, they offer a close approximation to the observed time series, which according to visual examination of the data appears to outperform the SCE-UA algorithm, at least during peak discharges. We further subject these results to a closer examination and also compare the flow duration curves that correspond to the BMU simulations of both training types to the SCE-optimized model and the observed discharges (Fig. 8). Within the range of the available Monte-Carlo output, the model realizations that correspond to the BMU of the SOM trained on Signature Indices (Fig. 8a) yield a rather close approximation to the flow exceedance probability characteristics of the measured data, with the exception of the low flows. The same holds true for the results obtained from the training on time series. The most notable differences regarding the flow duration curves affect the exceedance probability range between 10 and 20%, where the training on time series leads to better results, and all values below 2% exceedance probability, where the realizations obtained from the Signature Index trained SOM perform slightly better.

With respect to the BMU it is further insightful to inspect the quantization error of the observed data, calculated after Eq. (14), in order to get a rough measure on the quality of the BMU estimate: The average quantization error of the Signature Index map is 0.31, whereas the distance $\|x_c - m_c\|$ for

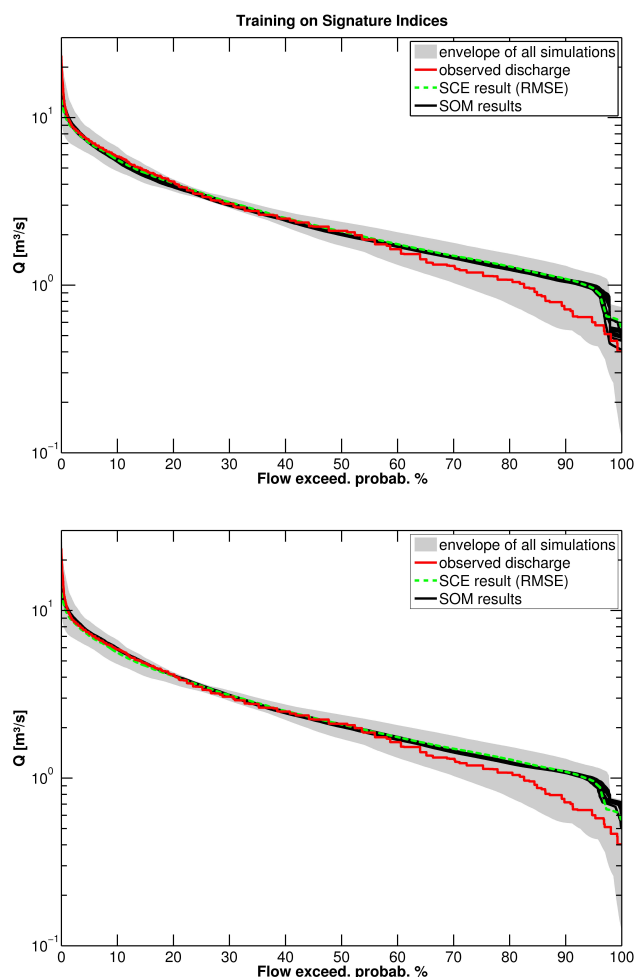**Fig. 8.** Comparison of flow duration curves for simulations attributed to the BMU.
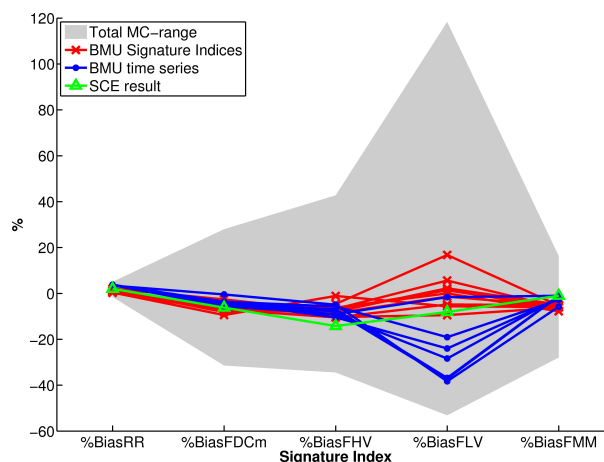

Fig. 9. Comparison of individual Signature Index ranges for the BMU-realizations corresponding to the SOM training on Signature Indices and the training on time series. Additionally, the Signature Indices obtained from minimizing the RMSE using the SCE-UA algorithm are displayed.

the BMU of the observed data (Qobs) on this map amounts to 1.1. This means that, in terms of the Signature Indices, the average distances between the model simulations and their corresponding BMU are not very much smaller compared to the dissimilarity between the reference vector of the BMU and the observed discharges. Consequently the observed discharge time series cannot be exceedingly different from the time series of the training data set and shows good correspondence with its BMU.

Comparing the Signature Index ranges of the model realizations projected to this BMU to the ranges of the entire Monte-Carlo data set (Fig. 9) the similarity of the simulations on this node, in terms of Signature Indices, as well as the compromise involved with minimizing the Signature Indices by identifying the BMU becomes visible. For all models the range of variation in overall water balance (%BiasRR) appears to be almost negligible when compared to the ranges of the remaining Signature Indices. The Signature Indices cal-

culated for the simulations attributed to the BMU of the SOM trained on time series, on the other hand, also show only moderate variation. Marked differences in the behaviour of these simulations, however, can be stated regarding the percent error in the FDC low-segment volume %BiasFLV. Here, the results obtained from the BMU of the SOM based on Signature Indices outperform the corresponding model realizations from the time series map. The result obtained by minimizing the RMSE using the SCE-UA algorithm displays Signature Index properties that are very similar to both of the aforementioned BMU realizations. Interestingly, the SCE-UA results perform slightly worse with respect to %BiasFHV and markedly better with respect to %BiasFLV, compared to most of the BMU realizations. This finding, however, does not fully coincide with the result suggested by the position of the RMSE optimum relative to the BMU: According to Fig. 6b and c the results for %BiasFLV are generally supposed to deteriorate towards the position of the RMSE optimum which lies approximately three map units above the BMU.

In Fig. 10 the normalized range of parameter values retrieved from the BMUs of the SOM trained on Signature Indices and the SOM trained on time series are shown. Here, it can be seen that in the first case identifying the BMU only leads to strongly constrained parameters regarding RetInf and maxInf, albeit their corresponding means are somewhat different. As expected, these findings correspond well to the arrangement of parameter values on the map in Fig. 5, where only RetInf and maxInf display an entirely ordered pattern. The parameters RetBasis and hL, on the other hand, only appear to be constrained for the realizations obtained from the time series trained SOM. This finding can most likely be explained with the parameter ranges being linked
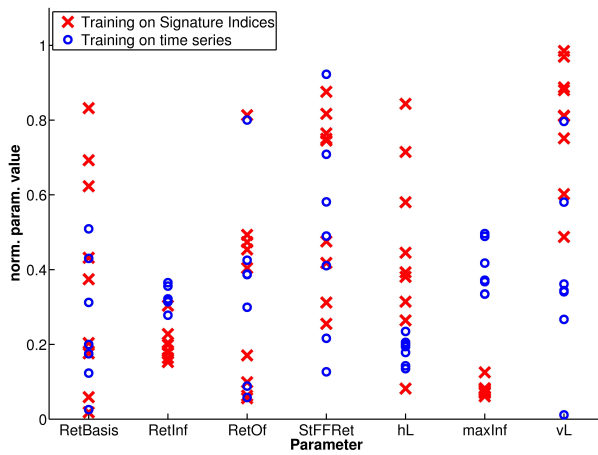
Fig. 10. Normalized ranges of parameter values retrieved from the BMUs of the SOM trained on Signature Indices and the SOM trained on time series.

to the location of the BMU, which in case of parameter hL coincides with a map region that lacks an orderly pattern with respect to the parameter values. The corresponding position of the BMU for the training on time series obviously falls into a map region where this is not the case and consequently its range in Fig. 10 is narrowly constrained.

4   Discussion and conclusions

The abovementioned results of our study are the product of two combined approaches: The Signature Measures serve to extract information from model time series data and, at the same time, considerably reduce the amount of data to be dealt with within the SOM algorithm. Gupta et al. (2008) and Yilmaz et al. (2008) argue the hydrological relevance and diagnostic value of such information for tracking inadequacies in watershed models and for guiding appropriate model improvements. By using the measures presented in Sect. 2.2, each time series is projected into a five-dimensional Signature Index Space. The underlying assumptions of our study, common to many multi-criteria evaluations, are that the Signature Indices are equally relevant and that the model is capable of reproducing the Signature Index spaces. The role of the Self-Organizing Map, in the second step of our approach, is to produce a discretized (and consequently data-compressed) mapping of the distribution in the Signature Index space onto a two-dimensional plane such that the patterns of hydrologically relevant information from a great number of model realizations can be conveyed in a comprehensive manner. In addition, the distribution of hydrologically relevant model properties is also linked to the corresponding parameter space that produced the model time series. The difference to the approach by Herbst and Casper (2008) lies in processing the raw time series data into informative indicators of hydrological model behaviour. Because in this step

the number of elements in each data item is reduced from several thousand time steps to 5 Signature Indices the dimensionality and thus the computational burden of training the Self-Organizing Map has been shrunk immensely such that 10 000 training cycles can now be completed in several minutes instead of requiring several days (all computations have been carried out on a 2×Dual-Core Opteron™ Server running at 2.59 GHz).

The high degree of correspondence in Signature patterns on both SOMs shown in Fig. 6 clearly suggests that the Signature Indices have been successfully extracted and have reproduced relevant parts of the information contained in the simulated time series. This, in a sense, points at a high degree of equivalence regarding the discriminatory power of the two SOM approaches and underscores the efficiency of using Signature Indices for SOM training on model output data. The comparison of Fig. 6a and b as well as Fig. 9 suggests that the SOM trained with time series data and the SOM based on Signature Indices most notably differ with respect to the representation of the long-term behaviour of the system. From the comparison of Fig. 6b and c it further becomes apparent that an SOM trained on the statistical performance measures used in Fig. 6c most probably would have extracted less independent "information" from the model data.

As some of the Signature Indices are derived from the flow exceedance probability curve it goes without saying that, despite their diagnostic value, not all Signature Indices are completely independent. On the other hand, it has been demonstrated that the way these measures covary for a given set of simulations can reveal much about the behaviour of the model. Most notably, model deficits and trade-offs in representing different watershed functions can immediately be visualized using adequate techniques for reproducing the SOM.

Herbst and Casper (2008) have already shown that preliminary information on parameter sensitivities can be provided by training a SOM on model output data. Using the SOM in conjunction with Signature Indices additionally allows us to interpret the function of individual model parameters in the context of hydrological theory even when completely ignoring the model structure.

The results from Sect. 3.1 have shown that the ability of a SOM to process and visualize multidimensional data can be exploited successfully for comprehending model behaviour in the hydrological context. Therefore, the SOM approach presented in this contribution also potentially constitutes a valuable tool for decision makers in as much as model realizations with specific Signature properties can be selected according to the purpose of the model application (an example is given in Fig. 4).

Furthermore, in a way analogous to Herbst and Casper (2008), the measured discharge time series has been projected onto the SOM. Because a model is generally not capable of perfectly reproducing all aspects of the observed data, a quantization error of 1.1 for the BMU of the measured

discharge time series appears to be acceptable compared to the average quantization error of the map. Note that a comparison between the average quantization errors achieved with the Signature Index SOM and the SOM trained on time series is not very meaningful due to extremely different dimensionalities of the input data spaces. The results given in Sect. 3.2 indicate that using the Signature Indices has positive effects on the way the SOM can be used to discriminate between different model realizations which consequently bring about improvements regarding the use of the SOM to identify the model results which most closely approximate a given time series. These improvements are probably attributable to the fact that using a set of Signature Indices for the SOM training (theoretically) attributes equal weights to the individual characteristics considered by them. When training on time series, however, the influence of particular events (e.g. high flow periods) on the SOM training is proportional to their frequency in the data. Thus, although only 4000 model simulations were available, the selection of the "optimal" models on the basis of a SOM trained on Signature Indices was sufficiently effective (and efficient) that, based on visual examination, the optimization using SCE-UA is outperformed. Of course, as only 4000 model realizations were used, the parameter space was quite sparsely sampled, and adding more data items to the training data set could easily help to refine these results. Because, in a general sense, a SOM can be understood as a means for capturing and analyzing multi-dimensional distributions it seems to be well suited for use in model identification and model uncertainty analysis. Based on the aforementioned results we suggest that the present SOM approach may constitute an initial step towards an alternative framework for model analysis and optimization. Further research could therefore, among other topics, be dedicated towards finding efficient re-sampling strategies to explore the parameter space. As to the aspect of using the SOM for multi-criteria optimization, it should be borne in mind that determining the BMU of the measured time series is equivalent to converting a multi-objective optimization to a single-criterion problem by means of weighting the objective function (Zadeh, 1963; Madsen, 2003). This might be of importance in cases where the multi-criteria Pareto front is non-convex; i.e. in general finding the BMU might not provide a genuine multi-criteria solution.

The Signature Indices used here by no means make a claim to completely cover all relevant information contained in the data and a series of other hydrological Signatures are easily imaginable, such as e.g. measures related to the representation of the snow cover or the height and timing of peak discharges. Just as little is known concerning the requirements of a "parsimonious" description of model behaviour/excluding redundant information on model behaviour. Therefore, further research on the development of SOM-based applications in the field of hydrological modelling will inevitably be confronted with the question of what constitutes a sufficient and informative metric/set of Signa-

ture Measures and the question concerning the number of neurons required to describe the behaviour of a model. This again leads to the more general problems of time series data mining. Defining and identifying the information content of the data constitutes the key towards a more meaningful and precise application of SOM and all other frameworks for the evaluation and analysis of hydrological models.

References

Abramowitz, G., Gupta, H. V., Pitman, A., Wang, Y., Leuning, R., Cleugh, H., and Hsu, K.-l.: Neural Error Regression Diagnosis (NERD): A Tool for Model Bias Identification and Prognostic Data Assimilation, J. Hydrometeorol., 7, 160–177, doi:10.1175/JHM479.1, 2006.

Abramowitz, G., Pitman, A., Gupta, H. V., Kowalczyk, E., and Wang, Y.: Systematic Bias in Land Surface Models, J. Hydrometeorol., 8, 989–1001, doi:10.1175/JHM628.1, 2007.

Abramowitz, G. and Gupta, H. V.: Toward a model space and model independence metric, Geophys. Res. Lett., 35, L05705, doi:10.1029/2007GL032834, 2008.

Abramowitz, G., Leuning, R., Clark, M., and Pitman, A.: Evaluating the Performance of Land Surface Models, J. Climate, 21, 5468–5481, doi:10.1175/2008JCLI2378.1, 2008.

Alhoniemi, E., Hollmén, J., Simula, O., and Vesanto, J.: Process Monitoring and Modeling using the Self-Organizing Map, Integr. Comput. Aid. E., 6, 3–14, 1999.

Ambroise, B., Perrin, J. L., and Reutenauer, D.: Multicriterion validation of a semidistributed conceptual model of the water cycle in the Fecht Catchment (Vosges Massif, France), Water Resour. Res., 31, 1467–1482, 1995.

Boogaard, H. F. P. v. d., Mynett, A. E., and Ali, M. S.: Self organizing feature maps for the analysis of hydrological and ecological data sets, in: Hydroinformatics '98, edited by: Babovic, V. M. and Larsen, L. C., Balkema, Rotterdam, The Netherlands, 733–740, 1998.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, Water Resour. Res., 36, 3663–3674, 2000.

Chang, D.-H.: Analysis and modeling of space-time organization of remotely sensed soil moisture, Department of Civil and Environmental Engineering, University of Cincinnati, Cincinnati, Ohio, USA, 169 pp., Ph. D. thesis, 2001.

Duan, Q., Sorooshian, S., and Gupta, V. K.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28, 1015–1031, doi:10.1029/91WR02985, 1992.

Franks, S. W., Gineste, P., Beven, K. J., and Merot, P.: On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process, Water Resour. Res., 34, 787–797, doi:10.1029/97WR03041, 1998.

Gallart, F., Latron, J., Llorens, P., and Beven, K.: Using internal catchment information to reduce the uncertainty of discharge and baseflow predictions, Adv. Water Resour., 30, 808–823, doi:10.1016/j.advwatres.2006.06.005, 2007.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour. Res., 34, 751–764, 1998.

Gupta, H. V., Sorooshian, S., Hogue, T. S., and Boyle, D. P.: Advances in Automatic Calibration of Watershed Models, in Calibration of Watershed Models, edited by Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application Series Vol. 6, AGU, Washington DC, USA, 9–28, 2003.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22, 3802–3813, doi:10.1002/hyp.6989, 2008.

Hall, M. J.: How well does your model fit the data? J. Hydroinform., 3, 49–55, 2001.

Haykin, S.: Neural networks - a comprehensive foundation, 2nd ed., New Jersey, USA, 842 pp., 1999.

Herbst, M. and Casper, M. C.: Towards model evaluation and identification using Self-Organizing Maps, Hydrol. Earth Syst. Sci., 12, 657–667, 2008,
http://www.hydrol-earth-syst-sci.net/12/657/2008/.

Hsu, K.-L., Gupta, H. V., Gao, X., Sorooshian, S., and Imam, B.: Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis, Water Resour. Res., 38, 1302, doi:10.1029/2001WR000795, 2002.

Huang, M., Liang, X., and Liang, Y.: A transferability study of model parameters for the variable infiltration capacity land surface scheme, J. Geophys. Res., 108(D22), 8864, doi:10.1029/2003JD003676, 2003.

Hydrotec: Rainfall-Runoff-Model NASIM – program documentation (in German), Hydrotec GmbH, Aachen, Germany, 579 pp., 2005.

Kalteh, A. M. and Berndtsson, R.: Interpolating monthly precipitation by self-organizing map (SOM) and multilayer perceptron (MLP), Hydrol. Sci. J., 52, 305–317, 2007.

Kaski, S.: Data Exploration Using Self Organizing Maps, Dr. thesis, Department of Computer Science and Engineering, Helsinki University of Technology, Helsinki, Finland, 57 pp., 1997.

Kohonen, T.: Self-Organizing Maps, 3rd ed., Information Sciences, Springer, Berlin, Heidelberg, New York, 501 pp., 2001.

Lamb, R., Beven, K., and Myrabø, S.: Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model, Adv. Water Resour., 22, 305–317, 1998.

Lane, S. N.: Assessment of rainfall-runoff models based upon wavelet analysis, Hydrol. Process., 21, 586–607, 2007.

Legates, D. R. and McCabe Jr., G. J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35, 233–241, doi:10.10291998WR900018, 1999.

Lin, G.-F. and Chen, L.-H.: Identification of homogeneous regions for regional frequency analysis using the self-organizing map, J. Hydrol., 324, 1–9, 2006.

Lin, Z. and Beck, M. B.: On the identification of model structure in hydrological and environmental systems, Water Resour. Res., 43, W02402, doi:10.1029/2005WR004796, 2007.

Lischeid, G.: A decision support system for mountain basin management using sparse data, EGU General Assembly 2006, Vienna, Geophys. Res. Abstr., 8, EGU06-A-04223, 2006.

Lourenço, F. C.: Exploratory Geospatial Data Analysis Using Self-Organizing Maps – Case Study of Portuguese Mainland Regions, M.Sc. thesis, Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa, Lissabon, Portugal, 143 pp., 2005.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Adv. Water Resour., 26, 205–216, 2003.

Mele, P. M. and Crowley, D. E.: Application of self-organizing maps for assessing soil biological quality, Agr. Ecosyst. Environ., 126, 139–152, 2008.

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

Peeters, L., Bação, F., Lobo, V., and Dassargues, A.: Exploratory data analysis and clustering of multivariate spatial hydrogeological data by means of GEO3DSOM, a variant of Kohonen's Self-Organizing Map, Hydrol. Earth Syst. Sci., 11, 1309–1321, 2007, http://www.hydrol-earth-syst-sci.net/11/1309/2007/.

Rajanayaka, C., Kulasiri, D., and Samarasinghe, S.: A Comparative Study of Parameter Estimation in Hydrology Modelling: Artificial Neural Networks and Curve Fitting Approaches, International Conference on Modelling and Simulation, MODSIM 2003 Congress, Townsville, Australia, 2003.

Reusser, D. E., Blume, T., Schaefli, B., and Zehe, E.: Analysing the temporal dynamics of model performance for hydrological models, Hydrol. Earth Syst. Sci. Discuss., 5, 3169–3211, 2008, http://www.hydrol-earth-syst-sci-discuss.net/5/3169/2008/.

Schaefli, B. and Gupta, H. V.: Do Nash values have value? Hydrol. Proc., 21, 2075–2080, 2007.

Schütze, N., Schmitz, G. H., and Petersohn, U.: Self-organizing maps with multiple input-output option for modeling the Richards equation and its inverse solution, Water Resour. Res., 41, W03022, doi:10.1029/2004WR003630, 2005.

Seibert, J.: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, Hydrol. Earth Syst. Sc., 4, 215–224, 2000.

Simon, G., Lendasse, A., Cottrell, M., Fort, J. C., and Verleysen, M.: Time series forecasting: Obtaining long term trends with self-organizing maps, Pattern Recogn. Lett., 26, 12, 1795–1808, doi:10.1016/j.patrec.2005.03.002, 2005.

Simula, O., Vesanto, J., Alhoniemi, E., and Hollmén, J.: Analysis and Modeling of Complex Systems Using the Self-Organizing Map, in: Neuro-Fuzzy Techniques for Intelligent Information Systems, edited by: Kasabov, N. and Kozma, R., Physica Verlag (Springer), 3–22, 1999.

Vesanto, J.: Using the SOM and Local Models in Time-Series Prediction, Workshop on Self-Organizing Maps, WSOM'97, Espoo, Finland, 1997.

Vesanto, J.: Using SOM in Data Mining, Licentiate's thesis,

Helsinki University of Technology, 57 pp., 2000a.

Vesanto, J.: Neural network tool for data mining: SOM toolbox, Symposium on Tool Environments and Development Methods for Intelligent Systems (TOOLMET2000), Oulu, Finnland, 184–196, 2000b.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J.: SOM Toolbox for Matlab 5, Helsinki University of Technology, Report A57, Espoo, Finland, 60 pp., 2000.

Vrugt, J. A., Weerts, A. H., and Bouten, W.: Information content of data foridentifying soil hydraulic parameters from outflow experiments, Soil Sci. Soc. Am. J., 65, 19–27, 2001.

Vrugt, J. A., Bouten, W., Gupta, H. V., and Sorooshian, S.: Toward improved identifiability of hydrologic model parameters: The information content of experimental data, Water Resour. Res., 38, doi:10.1029/2001WR001118, 2002.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., and Gupta, H. V.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis, Hydrol. Process., 17, 455–476, 2003a.

Wagener, T., Wheater, H. S., and Gupta, H. V.: Identification and Evaluation of Watershed Models, in Calibration of Watershed Models, edited by Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application Series Vol. 6, AGU, Washington DC, 29–47, 2003b.

Warren Liao, T.: Clustering of time series data – a survey, Pattern Recogn., 38, 1857–1874, doi:10.1016/j.patcog.2005.01.025, 2005.

Willmott, C. J.: On the validation of models, Phys. Geogr., 2, 184–194, 1981.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res., 44, W09417, doi:10.1029/2007WR006716, 2008.

Zadeh, L.: Optimality and non-scalar-valued performance criteria, IEEE T. Automat. Contr., 8, 59–60, 1963.

Appendix C

**Natural Hazards
and Earth
System Sciences**

Comparative analysis of model behaviour for flood prediction purposes using Self-Organizing Maps

M. Herbst[1]**, M. C. Casper**[1]**, J. Grundmann**[2]**, and O. Buchholz**[3]

[1]Department of Physical Geography, University of Trier, Germany
[2]Department of Hydrology and Meteorology, University of Dresden, Germany
[3]Hydrotec GmbH, Aachen, Germany

Abstract. Distributed watershed models constitute a key component in flood forecasting systems. It is widely recognized that models because of their structural differences have varying capabilities of capturing different aspects of the system behaviour equally well. Of course, this also applies to the reproduction of peak discharges by a simulation model which is of particular interest regarding the flood forecasting problem.

In our study we use a Self-Organizing Map (SOM) in combination with index measures which are derived from the flow duration curve in order to examine the conditions under which three different distributed watershed models are capable of reproducing flood events present in the calibration data. These indices are specifically conceptualized to extract data on the peak discharge characteristics of model output time series which are obtained from Monte-Carlo simulations with the distributed watershed models NASIM, LAR-SIM and WaSIM-ETH. The SOM helps to analyze this data by producing a discretized mapping of their distribution in the index space onto a two dimensional plane such that their pattern and consequently the patterns of model behaviour can be conveyed in a comprehensive manner. It is demonstrated how the SOM provides useful information about details of model behaviour and also helps identifying the model parameters that are relevant for the reproduction of peak discharges and thus for flood prediction problems. It is further shown how the SOM can be used to identify those parameter sets from among the Monte-Carlo data that most closely approximate the peak discharges of a measured time series. The results represent the characteristics of the observed time series with partially superior accuracy than the reference simula-

tion obtained by implementing a simple calibration strategy using the global optimization algorithm SCE-UA. The most prominent advantage of using SOM in the context of model analysis is that it allows to comparatively evaluating the data from two or more models. Our results highlight the individuality of the model realizations in terms of the index measures and shed a critical light on the use and implementation of simple and yet too rigorous calibration strategies.

1 Introduction

In the course of climate change the expected increase in the occurrence of meteorological conditions that trigger extreme flood events has raised the demand for operational flood management and flood forecasting systems, also in small- to medium-sized catchments (Kundzewicz et al., 2007; Merz and Didzun, 2005). A key component of these systems is very often represented by spatially distributed deterministic hydrological modelling systems whose properties and concepts have been subject to extensive research during the HORIX project. This project aims at developing an operational expert system for flood risk management in meso-scale watersheds considering prediction uncertainty (Disse et al., 2008) and forms part of the national research programme RIMAX (RIsk MAnagement of eXtrme flood events) which is dedicated to developing and implementing instruments towards improved flood risk management (Merz et al., 2007). An important aspect of the HORIX project is also to examine to what extent and under which circumstances different hydrological modelling systems support the prediction of (extreme) flood events in river catchments (Disse et al., 2007).

The discharge simulations that are produced using deterministic hydrological models are subject to different types of

Correspondence to: M. Herbst
(herbstm@uni-trier.de)

uncertainties which stem from the fact that every model is necessarily a conceptual and hence simplified representation of the natural system (e.g. Klemeš, 1983; Bossel, 2004; Sivapalan, 2005). As a consequence of this simplification, models are often not capable of covering the entire behavioural domain of the natural system with only one set of model parameters (Wagener et al., 2003). It is therefore recognized that models, because of their structural differences, have varying capabilities of capturing different aspects of the system behaviour equally well (Fenicia et al., 2007). In addition, the behavioural domain which can be reproduced by a model is further determined via calibration on historical discharge measurements which, in general, strives to account for the mean behaviour of the natural system although automatic calibration techniques (Duan et al., 2003) can emphasize, to some degree, different features of the data, depending on the performance measure which is chosen for the evaluation (Gupta et al., 1998). This, as a matter of course, excludes extreme events. Moore and Doherty (2005), however, have shown that the predictive capability of a model can be enhanced if weights are associated to those observations with the highest information content with respect to the required prediction. In order to maximize the probability that high discharge events can be simulated with a model it consequently appears reasonable to adapt the calibration strategy such that model performance in the domain of high discharges is emphasized.

In our study we use a Self-Organizing Map (SOM; Kohonen, 2001) in combination with index measures which are derived from the flow duration curve in order to examine the conditions under which three different distributed watershed models are capable of reproducing flood events present in the calibration data.

A Self-Organizing Map consists of an unsupervised learning neural network algorithm that performs a non-linear mapping of the dominant structures present in a high-dimensional data field onto a lower-dimensional grid. SOM has found almost countless applications in fields such as pattern recognition, image analysis (Kohonen, 2001) and exploratory data analysis (Kaski, 1997). However, applications related to hydrological modelling still seem to be the exception (see Minns and Hall, 2005). It has been used by Herbst et al. (2009) and Herbst and Casper (2008) for overall model evaluation and model identification purposes. Very recently, a SOM has been used by Reusser et al. (2008) to analyze the temporal dynamics of model behaviour. Kalteh et al. (2008) provide an overview of SOM applications in hydrological modelling.

In previous work in this field Herbst and Casper (2008) used the SOM to obtain a topologically ordered classification and clustering of the temporal patterns present in model outputs obtained from Monte-Carlo simulations. This clustering of entire time series allowed the authors to differentiate the spectrum of simulated time series with a high degree of discriminatory power and shows that the SOM can provide in-

sights into parameter sensitivities, while helping to constrain the model parameter space to the region that best represents the measured time series. The major shortcoming of this approach, however, was that, in the hydrological context, the underlying criteria of this mapping ("pattern") did not provide meaningful information on the trade-offs of model behaviour. In order to improve the extraction of information in terms of interpretable time series features (see also Boyle et al., 2000) Herbst et al. (2009) linked the SOM approach to the Signature Index concept by Gupta et al. (2008). Using the Signature Indices presented by Yilmaz et al. (2008), the dissimilarities between measured and simulated time series could now be expressed in hydrologically meaningful terms referring e.g. to water balance, mean runoff reaction velocity and the volume associated to long term base flow. Consequently, the SOM of these Signature Indices provided a concise summary of model behaviour which can potentially be used for model diagnostics. The present study follows a similar approach, however, with a more specific focus: The index measures we use to compare the simulated and the observed runoff were designed with the sole purpose of extracting different characteristics in the reproduction of peak flow and do not strictly follow the Signature Index concept by Gupta et al. (2008). A SOM of these indices is used to represent the spectra of model realizations obtained from Monte-Carlo simulations with the distributed watershed models NASIM (Hydrotec, 2005), LARSIM (Bremicker, 2000) and WaSIM-ETH (Schulla and Jasper, 2001) and subsequently analyze the individual trade-offs of model behaviour in the peak flow domain. It is demonstrated how the SOM of indices provide useful information about specific details of model behaviour and also helps identifying the model parameters that are relevant for the reproduction of peak discharges. It is further shown how the SOM can be used to identify those parameter sets from among the Monte-Carlo data that most closely approximate the peak discharges of a measured time series. At the first stage of this work (Sect. 3.1) the proposed technique is applied to each of the three models individually. At the second stage (Sect. 3.2) we directly compare the model realizations which were obtained from the three models with respect to the proposed criteria. The discriminatory power of the SOM is again used to identify those model realizations that most closely match the given set of criteria; however these realizations are selected from three different modelling systems. In order to assess to what extent constraints on the parameters contribute to enhancing the predictive capabilities of the three models to discharges that exceed the range of the calibration data, the parameter sets obtained from the SOM are applied to an extreme historical flood event which has not been part of the calibration data. The paper concludes with a discussion of the potential and the shortcomings of the presented approach (Sect. 4).

2 Methods and material

2.1 Models and data

In the present study we examine the results of 12 000 Monte-Carlo simulations of hourly discharge over a period of approximately two years. The time series were generated using the distributed watershed models NASIM, LARSIM and WaSIM-ETH, i.e. for each of the models we carried out 4000 simulations for the same test catchment based on input data for the period from 1 November 1994 to 28 October 1996. The test watershed (Fig. 1) is the 129 km^2 low-mountain range catchment "Schwarze Pockau" in Saxony (Germany), a tributary of the Freiberger Mulde (Elbe sub-basin) situated near the border to Czech Republic. The catchment extends from the ridges of the Erzgebirge (Ore Mountains) at approximately 980 m.a.s.l. northward to the runoff gaging station "Zöblitz" at 440 m.a.s.l. The mean discharge at this station is 2.31 m^3/s while the highest discharge ever measured was recorded on 13.08.2002 with 160 m^3/s. The return period for events of this magnitude is estimated to 200 a. About 40% of the catchment is covered with forest. The dominant soil type is a sandy loamy cambisol. The availability of discharge measurements from this catchment, especially during the extreme event of August 2002, render this catchment a good data source to investigate the capabilities of hydrological models of reproducing extreme discharges.

The rainfall data consists of spatially interpolated, hourly precipitation fields with a resolution of 1 km^2 which were generated based on daily measurements from three gaging stations and hourly measurements from one gaging station within the area (Fig. 1). Additionally, gaging stations from outside the test-catchment (not shown in Fig. 1) were included in order to assure proper conditions at the boundaries of the field. First, a two-dimensional external drift kriging (EDK 2D) is carried out on the daily measurements to get the estimate of the daily areal precipitation. In order to account for the temporal characteristics of the precipitation field additionally a separate EDK 2D is performed on the hourly precipitation measurements. Subsequently, the daily measurements are disaggregated according to the temporal distribution of the interpolated hourly precipitation. In both interpolations the square root of the elevation was used as drift parameter. EDK 2D was also applied in order to generate the spatio-temporal fields of wind speed, however, in this case elevation data determined the drift in a linear way. For the interpolation of global radiation and relative air humidity measurements a two-dimensional ordinary kriging was used. Streamflow was measured at the outlet of the catchment at gaging station "Zöblitz".

Because appropriate prior information on parameter distributions was missing the Monte-Carlo simulations were run using uniform random sampling. The corresponding parameter ranges as well as the fixed parameters were set based on prior knowledge acquired via manual expert calibration
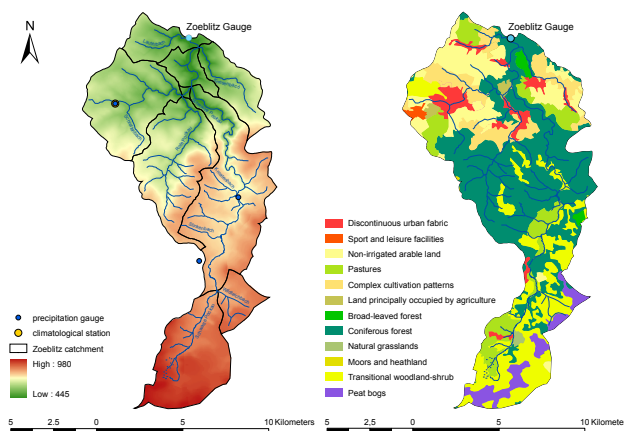

Fig. 1. The "Schwarze Pockau" low-mountain range catchment.

to the test watershed. It is assumed that these values represent the plausible parameter space for this watershed with very high probability. All parameters are related to the soil water balance and the vertical redistribution of flow components respectively. Parameters related to flood routing have not been considered for the Monte-Carlo simulation.

In the following, we give a brief outline of the model structures that were used for our study. Each of the models has found widespread application in different fields of hydrological modelling, including operational flood forecasting, throughout Germany and other countries. LARSIM and NASIM are distributed and operated commercially.

2.1.1 NASIM

NASIM (Hydrotec, 2005) is a conceptual distributed model that uses a spatial discretization based on sub-catchments. For the "Schwarze Pockau" watershed the pre-processing of spatial data resulted in 71 sub-catchments with a mean size of approximately 1.8 km^2. These are further subdivided into spatially homogeneous units with respect to soil and land use. Each of these elementary spatial units is again vertically divided into soil layers. All vertical processes that relate to soil and land use (soil moisture accounting, including interception, evapotranspiration, infiltration etc.) are calculated on the elementary unit scale. The resulting three lateral flow components are subsequently aggregated on the sub-cachment scale each passing an individual linear storage. Two of them, the interflow and surface flow, are in a prior step transformed by convolution with the time-area relationship to integrate sub catchment characteristics into the process of flow accumulation.

An outline of the principle elements of the model structure is given in Fig. 2. Note that the NASIM parameters examined in this study are unit less factors that modify the actual internal parameter values and act on the sub-catchment scale. The internal values are either based on global default values

Table 1. NASIM model parameters used for the Monte-Carlo simulation and their parameter (factor) ranges.

Name	Description	Factor Range	Internal Range	Unit
RetBasis	Storage coefficient factor for baseflow component	0.5–3.5	500[a]	[h]
RetInf	Storage coefficient factor for interflow component	2.0–6.0	50[a]	[h]
RetOf	Storage coefficient factor for surface runoff from unsealed surfaces	2.0–6.0	1.8–5.7[b]	[h]
StFFRet	Storage coefficient factor for surface runoff from urban areas	2.0–6.0	16[a]	[min]
hL	Horizontal hydraulic conductivity factor	2.0–8.0	1.5[a]	[mm]
maxInf	Maximum infiltration rate factor	0.025–1.025	11, 23, 30[c]	[mm/h]
vL	Vertical hydraulic conductivity factor	0.005–0.105	11, 23, 30[c]	[mm/h]

[a] fix value, not determined via pre-processing
[b] depeding on sub-catchment slope
[c] depeding on soil type
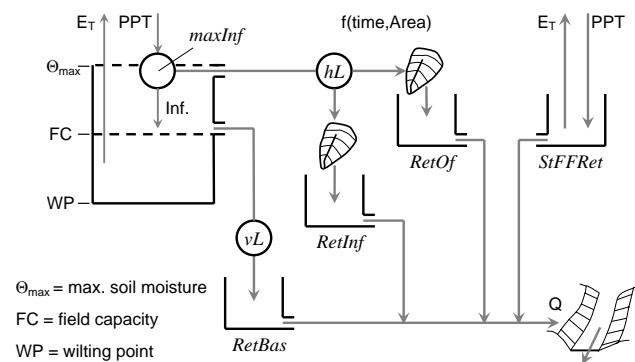
Fig. 2. Simplified schematic representation of the NASIM model structure; only those elements are reproduced that are considered in the scope of the present study. Parameters that have been subject to variation in the course of the Monte-Carlo simulation are printed in italic Times New Roman.

or have been determined individually for each sub-basin in the course of the spatial data pre-processing. The variation of these factors during the Monte-Carlo simulation, however, was performed with global values for all sub-catchments. Table 1 provides an overview of the calibration factors, internal values and their corresponding ranges. The parameter *maxInf* determines the maximum infiltration rate of the soil-moisture storage whereas the drainage is controlled by *vL*. The factors *RetOf*, *RetInf* and *RetBas* scale the storage coefficients for the quick, intermediate and the slow flow component, respectively. In the context of simulating flood events parameter *hL* potentially adopts a crucial role by determining the separation of exess flow into quick "overland flow" and intermediate "interflow" component. A special feature in NASIM is the representation of fast flow components from impervious urban areas whose retention is influenced by parameter *StFFRet*. However, in the Schwarze Pockau catchment only 6.6% of the area belongs to this land use type. Thus a dominant influence of this parameter is not expected. The internal

values modified by *RetOf* are determined in the course of the pre-processing depending on the slope in each sub-basin, while the internal *RetInf*, *RetBas*, *StFFRet* are set to global values. The correspondents of *maxInf* as well as well as *vL* are determined according to soil type. The ranges of calibration factors and internal parameter values of the Monte-Carlo simulation with NASIM are given in Table 1.

2.1.2 LARSIM

LARSIM (Bremicker, 2000) is operated using the same spatially distributed input data. However, in our study, a raster based spatial discretization with a resolution of 1 km was chosen. LARSIM considers coupled land use and soil compartments on a regular grid but does not explicitly account for the spatial distribution of soil and land use related field capacities on the sub-catchment scale. Instead, the amount of water which is allowed to infiltrate per time step is given as the difference between effective rainfall and overland flow. The sum of field capacity and air capacity yield the maximum soil moisture content. LARSIM then simulates the soil moisture balance using a variable contributing area function, similar to the approach implemented by the Xinanjiang model (Zhao, 1977): The proportion of contributing saturated areas is calculated as a function of mean soil water content and a conceptual parameter BSF (which is an exponent that controls the shape of the contributing saturated area function, see Fig. 3). The resulting total amount of saturated flow is subsequently partitioned into a quick and a slow sub-component, Q_{of} and Q_{of2}, depending on the threshold parameter $A2$. Discharge from lateral drainage Q_i ("interflow") as well as vertical percolation Q_b is represented using non-linear, empirical relationships such that essentially all flow components are controlled by the soil moisture storage and the actual soil moisture content: in Fig. 3 W_z denotes the minimum soil moisture content to generate interflow (it is considered a constant and set to 0.7 mm). The parameters D_{min} and D_{max} determine the minimum and maximum

amount of lateral drainage from the soil-moisture storage (in mm/d) which is further governed by the actual soil-moisture content W_0. Percolation into groundwater per time step Q_b is linearly controlled by parameter β. All flow components are subsequently forwarded to linear storage elements before they reach the river channel. The parameters EQD and $EQD2$ determine the storage coefficients that correspond to the sub-components of saturated flow Q_{of} and Q_{of2} respectively. They are linear scaling factors of the time of concentration in a sub-basin which is determined in the course of the pre-processing, i.e. they are proportional to the retention. For our study, the remaining storage coefficients are considered constant. The parameter ranges used to carry out the Monte-Carlo simulation are given in Table 2. The parameter values were identical for all sub-catchments during each run of the Monte-Carlo simulation.

2.1.3 WaSIM-ETH

WaSIM-ETH 6.4 version 2 (Schulla and Jasper, 1998) is operated with a raster based spatial discretization identical to the one used by LARSIM.

Infiltration of water into the soil is calculated for each grid cell following Peschke (1987). The remaining amount of water constitutes the surface flow component Q_d. Subsequently, soil water transport is simulated using the 1-D Richards differential equation on homogeneous soil columns which are determined by the spatial discretization. Soil hydraulic parameterization was carried out following the van Genuchten modelling scheme (Van Genuchten, 1976). The upper and lower boundary conditions are given by the amount of infiltrating water and the depth of the groundwater layer respectively. Lateral drainage Q_i results from the water balance calculations on the soil columns and is generated whenever the suction in the soil column falls below a given threshold ($\psi_m=3.45$ m). The drainage density parameter dr directly determines the amount of interflow which can be generated per time step. It expresses the drainage density of the (sub-)catchment as well as the anisotropy with regard to the vertical and horizontal hydraulic conductivities (Schulla and Jasper, 1998). For the simulations of our test catchment, however, no sub-basins were defined. A simple ground water model with a single linear storage approach is used to generate the slow discharge component Q_b. The subsequent concentration of the flow components is simulated using single linear storages and time-area functions on the catchment scale. The parameters kd and ki denote the storage coefficients of the surface runoff and the lateral flow, respectively. The resulting total discharge is calculated as the superposition of the flow components. A rough outline of the model is presented in Fig. 4. The parameter ranges for the Monte-Carlo simulation with WaSIM-ETH are reproduced in Table 3. Again, the parameter values remained identical for all sub-catchments during each simulation run.
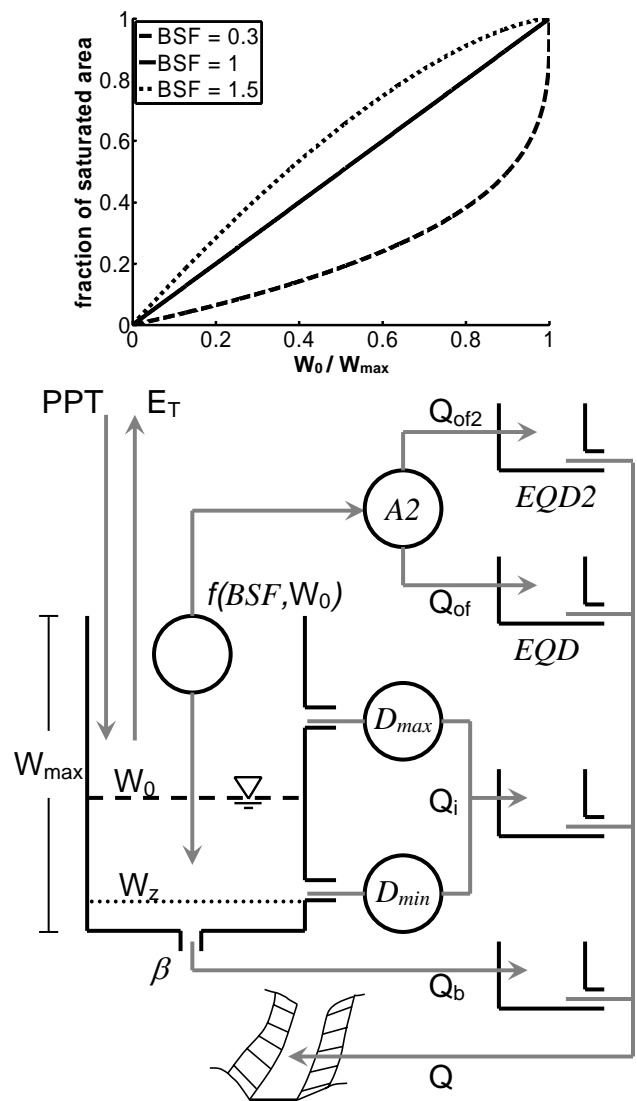

Fig. 3. Simplified schematic representation of the LARSIM model structure; only those elements are reproduced that are considered in the scope of the present study. Parameters that have been subject to variation in the course of the Monte-Carlo simulation are printed in italic Times New Roman.

As the focus of the present model evaluation lies on the reproduction of high discharges, the generation and concentration of flow through the model is considered to be the most important process here. Accordingly, the choice of model parameters for the Monte-Carlo simulation includes all parts of the particular model structures that seem to be most meaningful in this context. The resulting differences in the degrees of freedom between the models are considered here as an inevitable consequence of the particular model structure. In addition, it has to be taken into account that the number of available parameters can be strongly put into perspective by individual parameter sensitivities and by parameter interaction. In other words, model complexity is not a prerequisite

Table 2. LARSIM model parameters used for the Monte-Carlo simulation and their parameter ranges.

Name	Description	Unit	Range
EQD	Calibration factor for storage coefficient of fast runoff Q_{of}	[–]	100–5000
$EQD2$	Calibration factor for storage coefficient of fast runoff Q_{of2}	[–]	10–1000
BSF	Calibration factor of the "soil moisture" – saturated area function, variable contributing area approach	[–]	0.05–1.0
β	Drainage coefficient for deep storage	[1/d]	0.03–0.05
D_{min}	Minimum lateral drainage from soil storage	[mm/h]	0–5.0
D_{max}	Maximum lateral drainage from soil storage	[mm/h]	0–5.0
$A2$	Repartitioning factor for saturation overland flow and fast subsurface runoff	[mm/h]	0.8–3.0

Table 3. WaSIM-ETH model parameters used for the Monte-Carlo simulation and their parameter ranges.

Name	Description	Unit	Range
kd	Storage coefficient for surface runoff	[h]	0.1–40
ki	Storage coefficient for lateral flow	[h]	0.1–100
dr	Drainage density/anisotropy parameter	[1/m]	0.5–100

for good model performance (see e.g. Gan and Biftu, 2003). Thus, we see no strong reason to assume that a model would have less capabilities of reproducing certain runoff characteristics due to its degrees of freedom, even more as the present study focuses on a very specific aspect of model behaviour.

2.2 Derivation of index measures from the flow duration curve

In order to capture information on different characteristics of model behaviour within a specific domain of flow response we follow an approach which is adapted from the work of Gupta et al. (2008) and Yilmaz et al. (2008): Five index measures are derived based on the evaluation of simulated and observed flow duration curve properties. In contrast to commonly used statistical objective functions (e.g. see Legates and McCabe Jr., 1999) the "Signature Indices" presented by Yilmaz et al. (2008) constitute hydrologically meaningful measures of system response. In this respect, the indices we use differ from the concept proposed by Gupta et al. (2008) insofar as their diagnostic relation to different elements of the model structure as well as to the natural system is less obvious. In order to analyze the reproduction of flood events in detail the indices were conceptualized to focus solely on the characteristics of discharge events with an exceedance probability below a given threshold which is derived from the flow duration curve (Fig. 5). In our study this specific threshold is determined by visual examination of the slope of the observed flow duration curve which, in our example, shows a marked increase at 2%. The remaining section of the flow duration curve is further subdivided at 0.42%, following the
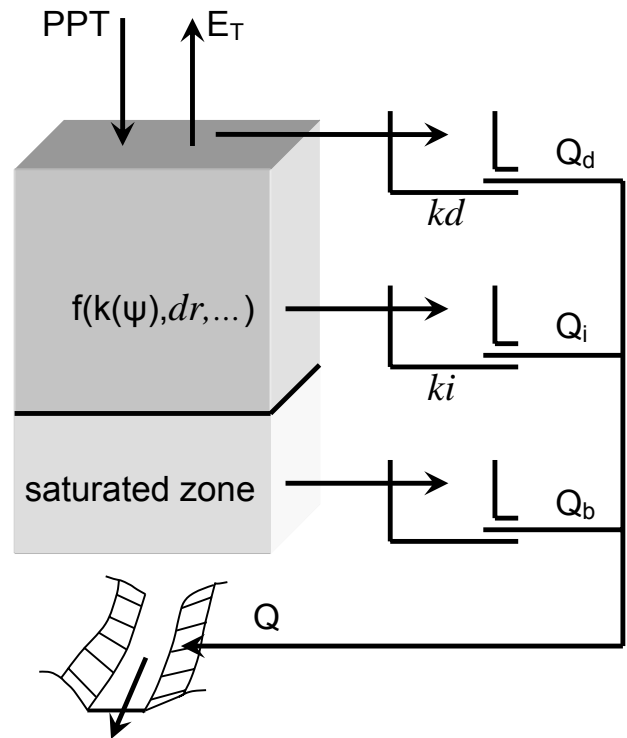

Fig. 4. Simplified schematic representation of the WaSIM-ETH model structure – only those elements are reproduced that are considered in the scope of the present study. Parameters that have been subject to variation in the course of the Monte-Carlo simulation are printed in italic Times New Roman.

same approach (Fig. 5). For each of these subsections individual index measures are calculated according to Eqs. (1) and (2). According to Yilmaz et al. (2008), the percent difference in slope of a flow duration curve segment relative to the observations is given as

$$\%BiasFDC =$$
$$\frac{\left(\log(Q\text{sim}_i) - \log(Q\text{sim}_j)\right) - \left(\log(Q\text{obs}_i) - \log(Q\text{obs}_j)\right)}{\left(\log(Q\text{obs}_i) - \log(Q\text{obs}_j)\right)} \cdot 100 \quad (1)$$
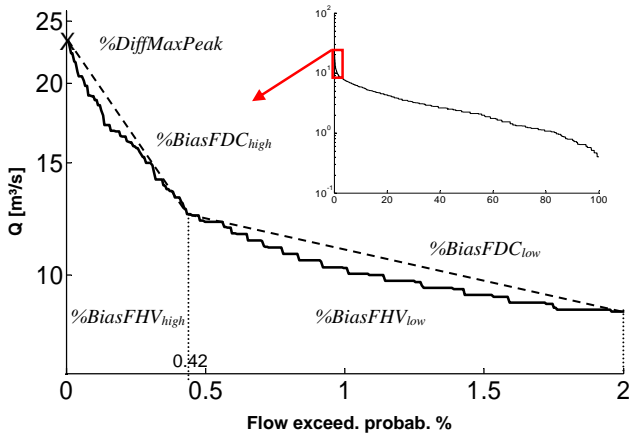
Fig. 5. Derivation of index measures from the upper 2% section of the flow duration curve (FDC).

Fig. 6. The iterative training process of SOM (Herbst and Casper, 2008).

where i and j denote the thresholds that define a segment of the flow duration curve; Qsim being the simulated discharges and Qobs being the corresponding observations. Given the observed flow duration curve in Fig. 5 we define the slope of the lower section of the flow duration curve segment *%BiasFDC*$_{low}$ as *%BiasFDC* with $i=2$ and $j=0.42$. Accordingly, the slope of the upper section of the flow duration curve segment *%BiasFDC*$_{high}$ is defined as *%BiasFDC* with $i=0.42$ and $j=0$. Further, the percentage of bias in the flow duration curve high volume segment is calculated based on Yilmaz et al. (2008) as

$$\%BiasFHV = \frac{\sum\limits_{h}(Q\text{sim}_h - Q\text{obs}_h)}{\sum\limits_{h}(Q\text{obs}_h)} \cdot 100 \qquad (2)$$

where h denotes the index of all discharge values with exceedance probabilities higher than i and lower than j. Again, we define the bias for the lower flow duration curve segment volume *%BiasFHV*$_{low}$ as *%BiasFHV* with $i=2$ and $j=0.42$. Correspondingly, we define *%BiasFHV*$_{high}$ as *%BiasFHV* with $i=0.42$ and $j=0$. In addition, the percentage of error in maximum peak discharge *%DiffMaxPeak* is determined after Eq. (3).

$$\%DiffMaxPeak = \frac{Q\text{sim}_H - Q\text{obs}_H}{Q\text{obs}_H} \cdot 100 \qquad (3)$$

with the index number of the highest element of the flow duration curve being H.

As none of the model parameters that are subject to variation in the Monte-Carlo simulation is related to flood routing or exerts a significant influence on the timing of the discharge peaks we refrained from examining potential time lags between the simulated data and the observations. However, in a more general model evaluation problem, this might be a recommendable procedure.
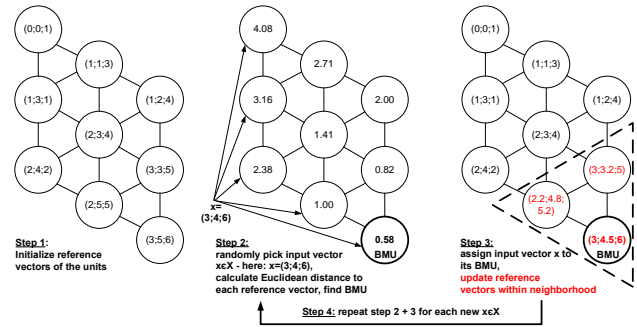
2.3 Self-organizing maps

SOM is an unsupervised learning neural network algorithm that is applied to high-dimensional data sets in order to categorize the range of data patterns that occur in it and to extract a set of characteristics that describe its multidimensional distribution. A SOM essentially performs a non-linear mapping of vectorial input data items onto a discrete, low-dimensional grid. Most commonly a two-dimensional, rectangular grid with hexagonal topology is used. In contrast to common Vector Quantization methods or k-Means clustering, SOM is topology preserving, i.e. nearby locations on this mapping are attributed to similar data patterns. Likewise, the distance between two nodes on the mapping is proportional to the dissimilarity of the data items they represent. Each input data item $\boldsymbol{x}$ of the training data set X that has to be examined is considered as a vector $x=[x_1, x_2, \ldots, x_n]^T \in \Re^n$, with n being the dimension of the input data space. Let X represent a set of index vectors calculated according to Sect. 2.2, thus $n=5$. A SOM consists of a fixed number of k neurons that are arranged on a regular grid whose dimensions can be determined by means of heuristic algorithms, if no other preferences are made. Throughout this paper the terms "neuron", "node" and "map unit" are used synonymously. Figure 6 provides a schematic representation of the process of self-organization which, in the following, is explained based on the paper by Herbst et al. (2009).:

Each neuron i is represented through a reference vector

$$m_i = [\mu_{i1}, \mu_{i2}, \ldots, \mu_{in}]^T \in \Re^n \qquad (4)$$

whose dimension n equals the number of elements in an input data vector $\boldsymbol{x} \in X$. Typically, the reference vectors $\boldsymbol{m}_i$ are initialized to small random values. However, in order to assure faster and more reliable convergence of the map, we initialize the $\boldsymbol{m}_i$ along the two greatest principal component eigenvectors of the data (Kohonen, 2001). In the classic sequential training the SOM is trained iteratively: In the first

step an input data item $x \in X$ is randomly selected and the Euclidean distance

$$d_i = \sqrt{\sum_{j=1}^{n} (x_j - m_{ij})^2} \quad i = 1 \dots k; \ j = 1 \dots n \qquad (5)$$

between x and each reference vector m_i is computed (theoretically any appropriate metric can be used as a measure of similarity). The "winning neuron" (also called the best-matching unit BMU of x) is the map element c whose reference vector m_c has the smallest distance d_c to x with

$$d_c = \min_i \{\|x - m_i\|\} . \qquad (6)$$

In the next step the reference vector m_i and all of its neighbouring neurons are updated according to

$$m_i (t+1) = m_i (t) + \alpha (t) h_{ci} (t) [x (t) - m_i (t)] \qquad (7)$$

where $m_i(t)$ is the current weight vector at iteration step t. Thus, the rate of change for each node of the map is scaled by three factors: a) the difference $(x(t) - m_i(t))$ between the input data set x and the prototype vector m_i b) the size of a neighbourhood function h_{ci} which decreases monotonically to zero with t and with distance from the winning neuron and c) a learning rate factor $\alpha(t)$ which gradually lowers the height of the neighbourhood function as the iteration advances. For h_{ci} it is common to use the Gaussian function

$$h_{ci} (t) = \exp \left(-\frac{\|r_c - r_i\|^2}{2\sigma^2 (t)} \right) \qquad (8)$$

where $\sigma(t)$ defines the width of the topological neighbourhood, and both $\sigma(t)$ and $\alpha(t)$ decrease monotonically with t. Note that an exact choice of the function $\alpha(t)$ is not required (Kohonen, 2001). Repeated cycling through the training steps causes different nodes and regions of the map to be "tuned" to specific domains of the input space. Importantly, the enforced local interaction between the SOM nodes results in the map gradually developing an ordered and smooth representation of the input data space (Kaski, 1997).

In this work, however, we used Kohonen's "batch-training" algorithm (Vesanto, 2000) to speed up the training process. Here, in each training step the data set is partitioned according to the Voronoi regions of the m_i. Instead of sequentially running through all data items in each training cycle the whole data set X is presented to the map as a whole at each training cycle. The reference vectors are updated according to the weighted average of the data samples

$$m_i (t+1) = \frac{\sum_{l=1}^{N} h_{ci}(t) x_l}{\sum_{l=1}^{N} h_{ci}(t)} \qquad (9)$$

where c is the index number of the BMU of data set x_l, and N is the number of data samples. This variant of the training does not make use of the learning rate factor $\alpha(t)$.

In the course of the training the reference vectors are "tuned" to the different patterns contained in the input data. The final reference vectors form a discrete approximation of the input data distribution. Thus, patterns that occur more frequently in the input space are mapped onto a larger area. Note that, as the number of neurons – and consequently the number of reference vectors – is much smaller than the number of data items used for the training, SOM can also be seen as a data compression method.

In our study we also make use of the fact that, once its training is finished, the SOM can be applied to project an input data vector y onto the map which has not been part of the training data set. This means that according to Eq. (6) the neuron $c(y)$ with reference vector $m_{c(y)}$ is determined for which

$$\|y - m_{c(y)}\| = \min_i \{\|y - m_i\|\} . \qquad (10)$$

Neuron $c(y)$ then represents the domain of input data patterns from X that is most similar to y. It follows that the set of data items $\hat{X} \subset X$ which is attributed to $c(y)$ represents those training data items that are most similar to y with respect to the criterion given by Eqs. (5) and (10). The neuron $c(y)$ is called the "best-matching unit" (BMU) of y.

2.4 Data preparation and training of the SOM

For each of the 4000 time series obtained by running a Monte-Carlo simulation (Sect. 2.1) a set of five index measures was calculated according to Eqs. (1–3) (Sect. 2.2). The procedure was carried out for each of the three models.

Prior to the SOM training, each index was normalized to a value having zero mean and variance of one using a linear transformation such that high index values do not exert a disproportionate influence on the training. The side lengths of the map as well as the initial reference vectors were determined by means of a heuristic algorithm involving the calculation of the two biggest eigenvalues of the covariance matrix of the data (Vesanto et al., 2000). For more details on the data preparation and the training please see Herbst et al. (2009).

At the first stage of our study the three data sets were treated individually. Subsequently, the data preparation and the training were repeated with the combined data set of all three models.

For the SOM training as well as for a part of the evaluation procedures the "SOM Toolbox for MATLAB™" (Helsinki University of Technology, http://www.cis.hut.fi/projects/somtoolbox/) was used.

2.5 Evaluation of SOM results

Generally, the number of neurons on the maps is much smaller than the number of data sets used for the training. As

a consequence of this, every neuron represents a set of simulation runs and their respective index value pattern. In the following, we evaluated the index properties of the individual nodes by de-normalizing the reference vectors of the maps. Each of the nodes/reference vectors represent the mean index value properties of a small sub-set of model data used for the training. In the following, these reference vector index values are visualized by means of a small, coloured bar plot for each node. In a bar plot visualization the position of the BMU can easily be identified by the map unit with the "flattest" bars. Note that the height of the individual bars is scaled relative to the range of the corresponding index. Also colour coding of the index values is used, whereas the same map grid is reproduced five times with a colouring corresponding to the distribution of the individual index values (so-called component planes).

As a result of the self-organizing process that takes place in the course of the training, the data items which are grouped to such a sub-set have similar properties with regard to their five index values. Due to the topological properties of the mapping the distance between two nodes on the map is roughly a function of the dissimilarities between the data sets attributed to these nodes. Please note that, to some extent, the SOM embodies statistical properties, e.g. the number of reference vectors on a map that display a certain type of quality is proportional to the number of data sets with that property. As a simple measure of the quality of the mapping the "quantization error" (Kohonen, 2001) $\bar{d}$ is calculated using Eq. (10).

$$\bar{d} = \frac{1}{N} \sum_{p=1}^{N} \left\| \boldsymbol{x}_p - \boldsymbol{m}_{c(p)} \right\| \qquad (11)$$

It represents the average distance of each data vector $\boldsymbol{x}_p$ of the N input data items contained in the training data X to its associated BMU reference vector $\boldsymbol{m}_{c(p)}$, with p being the index of the data items (not to be confused with the index values of Sect. 2.2!).

We further take advantage of the possibility to label the input data items that are attributed to each neuron via the training. That way, each input data item is linked to a model parameter set and its original simulated time series. Thus, the neurons of the map can be evaluated with respect to the model parameters, e.g. by calculating the mean values of each parameter for the individual map units. The distribution of parameter values over the map is again visualized by means of colour coding. In doing so, the same map grid is reproduced for each parameter, however, with different colouring according to the distribution of parameter values. In the following, this type of visualization is referred to as parameter plane. Corresponding patterns on a component plane and parameter plane indicate that an index value is governed to a large extent by a particular parameter. Moreover, an irregular pattern on the parameter plane is indicative of parameter

insensitivity, according to the components which were used to train the map.

For each map grid (i.e. for each model) the BMU of the measured discharge time series is determined according to Eq. (10). Following Sect. 2.2 (Eqs. 1–3) the time series of observed discharges Qobs maps as $\boldsymbol{y}=[0\ 0\ 0\ 0\ 0]^T$ into the index space. We then calculate the quantization error of the BMU

$$\bar{d}_{\text{BMU}} = \frac{1}{n} \sum_{r=1}^{\hat{N}} \left\| \hat{\boldsymbol{x}}_r - \boldsymbol{m}_{c(y)} \right\| \qquad (12)$$

in order to obtain a rough indicator of how close the data items $\hat{\boldsymbol{x}}_r \in \hat{X}$ with $r = 1 \ldots \hat{N}$ which are attributed to the BMU $c(\boldsymbol{y})$ of the observation (Eq. 10) approximate the observation (represented by $\boldsymbol{y}=[0\ 0\ 0\ 0\ 0]^T$). In Eq. 11 $\hat{N}$ denotes the number of data items in $\hat{X}$. Note, that it is possible to identify a BMU for any data set that has the same dimensionality as the input data, irrespective of its distance from the observations.

The data items on the BMU $c(\boldsymbol{y})$ of a map also correspond to one or more model parameter sets which are subsequently used to simulate an extreme flood event from August 2002 that has not been part of the Monte-Carlo data set. As a reference, we visually compare these simulations to the results we obtained by using the shuffled complex evolution optimization algorithm (SCE-UA, Duan et al., 1992) to find a parameter set that minimizes the root of the mean squared error (RMSE) for the same period of time for which a simple, model specific, weighting scheme after Casper et al. (2009) is used. This scheme basically applies a higher weight to all time steps with a discharge higher than three times the mean discharge ($Q>$3MQ).

3 Results

3.1 SOMs generated from the individual model data sets

Although from each of the models the same amount of data items was processed, the maps for NASIM, LARSIM and WaSIM-ETH slightly differ in number of neurons and side lengths which is caused by the initialization of the maps according to Sect. 2.4. The overall quantization error $\bar{d}$ (Eq. 11) for the three mappings ranges between 0.236 (NASIM) and 0.278 (WaSIM-ETH). Thus, it can be assumed that the SOM provides a good fit of the model data. The distributions of reference vector properties over the map support that the model data has been arranged by similarity over the maps. No void, i.e. interpolative, units are present. In the following, an account of the results for the individual models is given. Note that, as to the simulated time series illustrations, only a representative period of the entire simulated time series is reproduced in the following in order to assure better readability of the figures. Representations of the flow duration curve
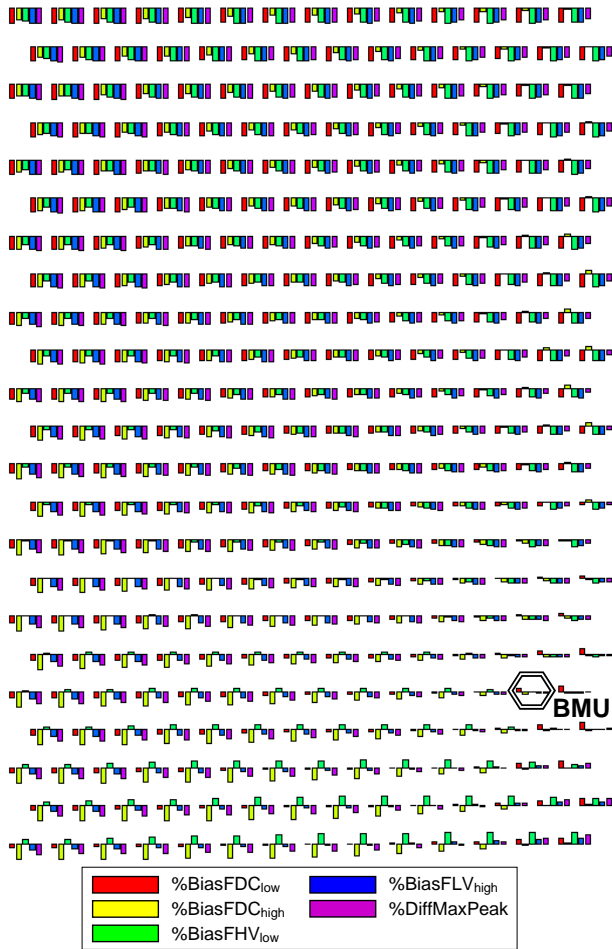
Fig. 7. NASIM: Distribution of index properties on the map displayed as bar plots and the position of the best-matching unit (BMU). Note that the bars have individual relative scales.
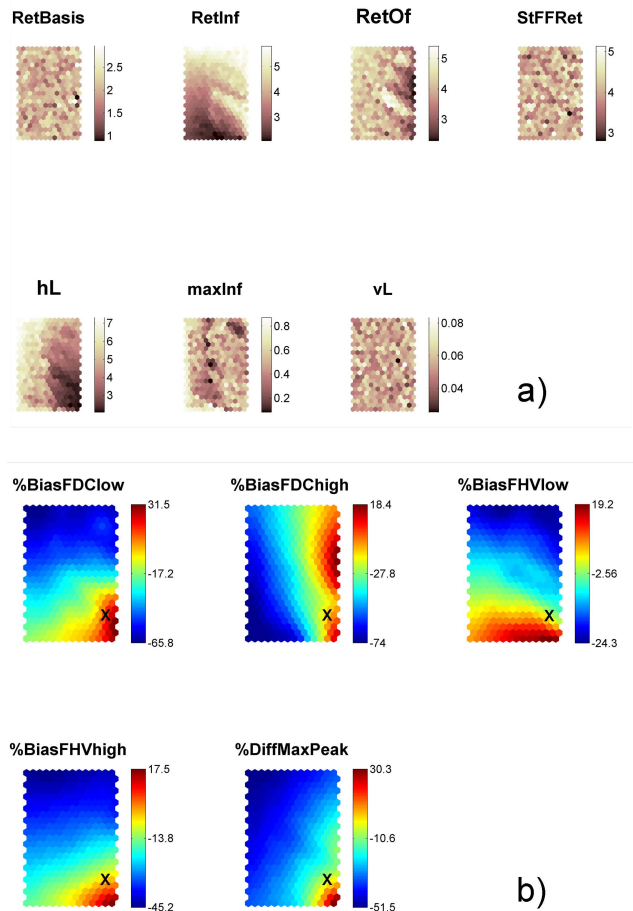

Fig. 8. NASIM: **(a)** Parameter plane, i.e. mean values of each model parameter for the simulations projected onto the individual map elements. **(b)** Distribution of reference vector (i.e. indices) properties on the map. The position of the BMU is marked.

(FDC), however, always refer to the full length of simulated discharge time series.

3.1.1 NASIM

The bar plot (Fig. 7) reveals that a significant proportion of simulation runs which were attributed to the upper half of the map underestimated all five properties that are represented by the indices. The increase of $\%BiasFDC_{high}$ from the left to the right hand side implies a general increase in peak runoff reaction in this direction of the map. The remaining index components generally tend to smaller negative or positive values towards the lower part and the right hand side of the map. From the values of $\%DiffMaxPeak$ in Fig. 7 it immediately becomes obvious that only very few simulations exceeded the measured peak discharge.

A comparison of the parameter plane Fig. 8a with the component plane Fig. 8b indicates that the increase in peak runoff reaction $\%BiasFDC_{high}$ is largely influenced by parameter

hL in combination with the parameter for the retention of "overland flow", $RetOF$. Figure 8a also reflects that parameter $RetOf$ remains insensitive with respect to the indices as long as hL has high values. According to Sect. 2.1, this behaviour is evident because the generation of "overland flow" is overridden by hL. Moreover, $RetInf$, which governs the retention of water allocated to interflow, with high probability exerts an influence on the increase in $\%BiasFHV_{low}$, which consequently provides a rather simple explanation for the position of the BMU on the map. Further, it can be seen from Fig. 8a that parameter $maxInf$ is – at best – only partially sensitive. vL, $RetBasis$ and $StFFRet$ are insensitive here because they are linked to the generation of "base flow" and runoff from urban or impervious areas which for the "Schwarze Pockau" catchment only comprise 6.6% of the total area.

Correspondences in the distribution patterns in Fig. 8b reveal significant correlations between the indices with respect to the behaviour of NASIM. However, each component plane is scaled separately. Therefore, the individual optima of the
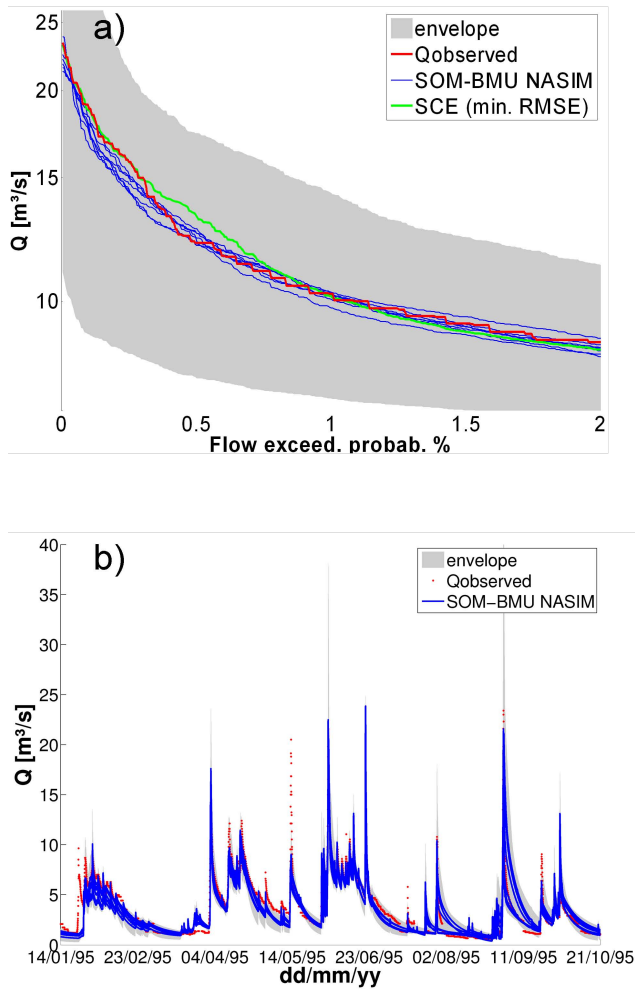
Fig. 10. NASIM: Results for the validation event August 2002 (BMU realizations and SCE-UA, RMSE).

Fig. 9. NASIM: **(a)** Flow duration curves (upper 2% section): Simulations corresponding to the BMU compared to the observed discharge and the results of an optimization approach using SCE-UA. **(b)** Time series corresponding to the BMU compared to the observed discharge.

indices still do not coincide on the same map location. The scales of the component planes additionally give insight into the true lengths of the bar plots in Fig. 7.

The quantization error for the BMU of the observed discharge time series, calculated according to Eq. (11), yields $\bar{d}_{BMU}=0.85$, which suggests that the observations are located somewhat off the model data obtained from the Monte-Carlo Simulation. Nevertheless, from the simulation results in Fig. 9a it can be seen that the parameter sets that were retrieved from the BMU associated to the observed runoff reproduce the characteristics of the measured FDC with very high accuracy. However, this requirement is also satisfied surprisingly well by the reference simulation which was obtained using the SCE-UA algorithm in combination with a simple weighting scheme. Notwithstanding, the time series of simulated discharge (Fig. 9b) shows that, with the excep-
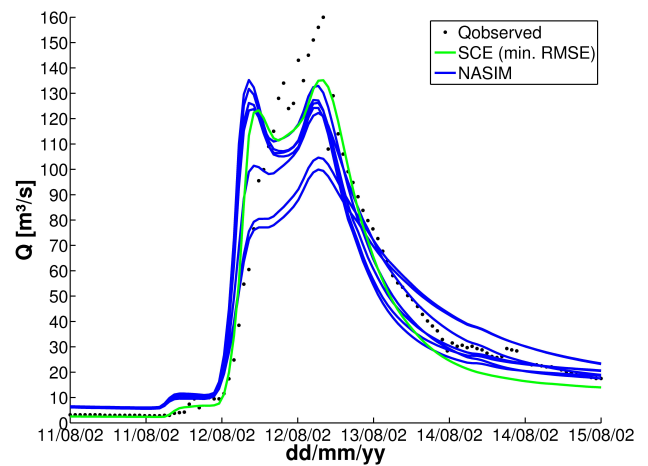
tion of only a few events, the runoff peaks could be reproduced well. As the FDCs grow almost congruent towards the ordinate it does not surprise that a part of the simulation runs attributed to the BMU performs equally well during the validation event compared to the SCE-UA reference simulation, yet none of the model realizations is able to reach the peak flow (Fig. 10).

3.1.2 LARSIM

Figure 11 shows that LARSIM, contrary to NASIM, tends to overestimate almost the entire set of indices whereas for at least the lower third of the map this tendency is very pronounced. Lower or negative index values as well as sporadic underestimation of runoff reaction (*%BiasFDC*$_{high}$) and – volume (*%BiasFHV*$_{low}$), are largely recorded in the upper regions. The only index, however, which is constantly overestimated throughout the data set is the runoff reaction during all time steps corresponding to the lower section of the FDC, *%BiasFDC*$_{low}$. The maximum peak discharge, expressed by *%DiffMaxPeak*, is also overestimated throughout large portions of the model data.

Towards the right hand corners a marked increase in *%BiasFDC*$_{high}$ is superimposed to the comparatively monotonous pattern of index combinations. From Fig. 12a it can be seen that parameter *EQD2* decreases in the same direction, which most likely reveals a main control for the runoff reaction in the fastest portion of flow. Likewise, the error in maximum peak flow, which is expressed by *%DiffPeakMax,* grows strongly positive towards the lower right hand corner. The volume allocated to flow corresponding to the lower branch of the FDC (*%BiasFHV*$_{low}$), and partially also *%BiasFHV*$_{high}$, increases towards the lower left corner of the map (Fig. 12b) which, according to Fig. 12a, indicates that these features are likely to be governed by parameter
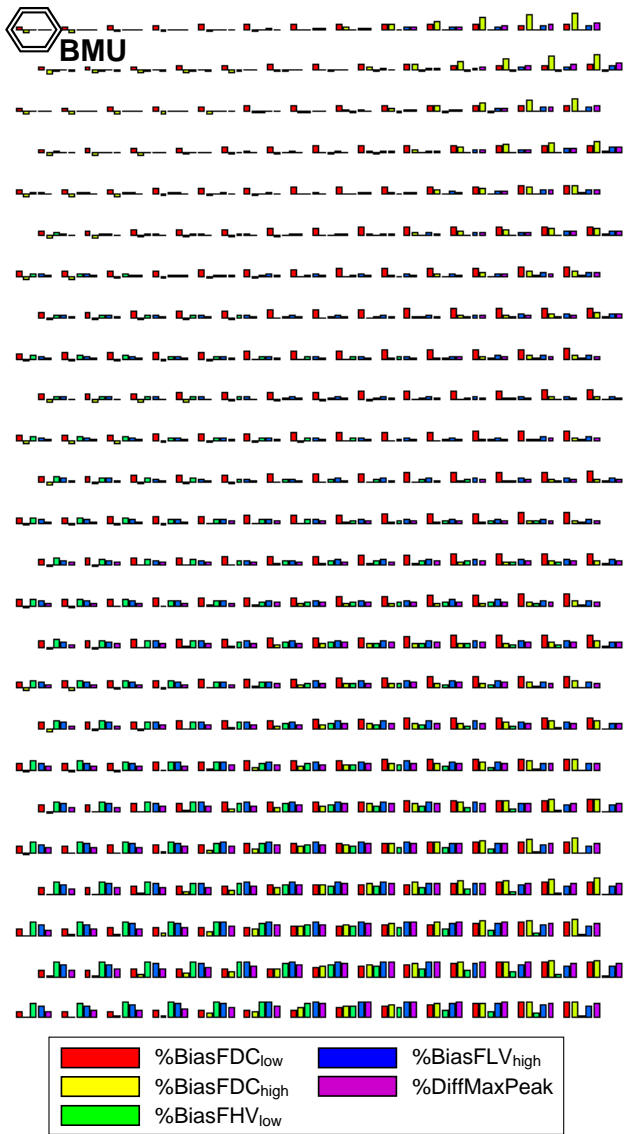
Fig. 11. LARSIM: Distribution of index properties on the map displayed as bar plots and the position of the best-matching unit (BMU). Note that the bars have individual relative scales.

Fig. 12. LARSIM: **(a)** Parameter plane, i.e. mean values of each model parameter for the simulations projected onto the individual map elements. **(b)** Distribution of reference vector (i.e. indices) properties on the map. The position of the BMU is marked (top left corner of the map.

EQD. As to the remaining parameters, no further correlations with indices are clearly apparent, although at least parameter *A2* shows a marked sensitivity with respect to data sets allocated to the right hand side of the mapping which points to some degree to an interaction between *EQD* and/or *EQD2* as well as *A2*.

The influence of *EQD2* on *%BiasFDC*$_{high}$ can easily be explained following Sect. 2.1: As *EQD2* is a scaling factor for the retention of the "fast" saturated flow component Q_{of2}, it largely controls the volume per time step which is allocated to peak discharges. Its sensitivity, however, depends on the threshold parameter *A2* which at the same time governs the function of storage coefficient *EQD*. As documented by the
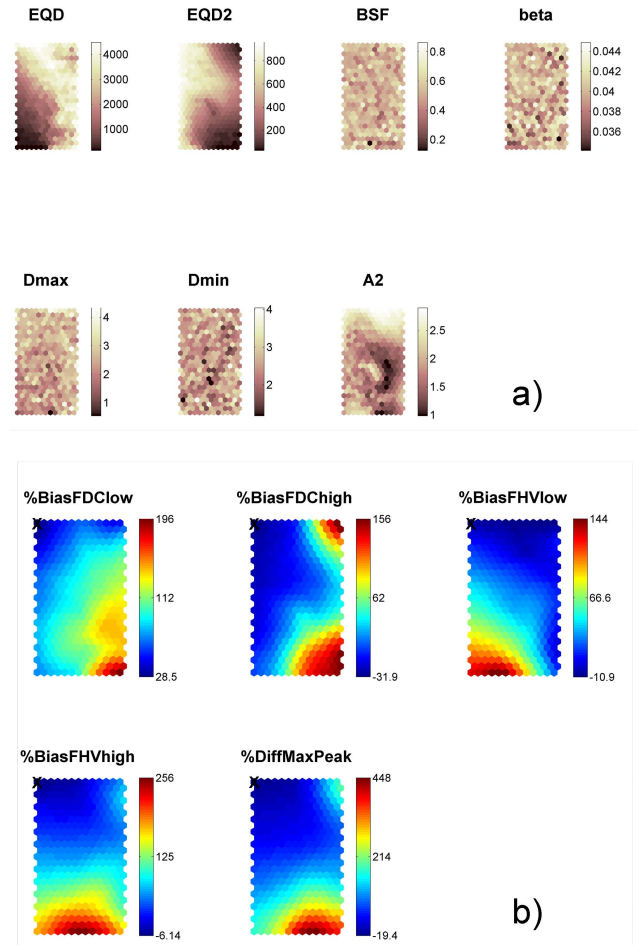
corresponding scales of the component planes in Fig. 12b, the sensitivity of *%DiffMaxPeak* and *%BiasFDC*$_{high}$ is extremely high. But also the volumes *%BiasFHV*$_{low}$ and *%BiasFHV*$_{high}$ seem to react with sharp gradients on changes of *EQD* and *EQD2*. Following Sect. 2.1 it does not surprise that the "base flow" storage coefficient β as well as D_{min} or even D_{max} do not show any apparent sensitivity according to Fig. 12a. However, the lack of sensitivity with respect to parameter *BSF*, which controls the generation of saturated flow via a variable contributing area approach, is unexpected and does not lend itself to a straightforward explanation.

The impression that LARSIM does not seem to be capable of simultaneously meeting the constraints imposed by the five indices is already conveyed by Fig. 11. The comparatively high quantization error for the BMU of the observed discharge time series $\bar{d}_{BMU}=1.1$ and the fact that this BMU is located on an extremely marginal position in the upper left
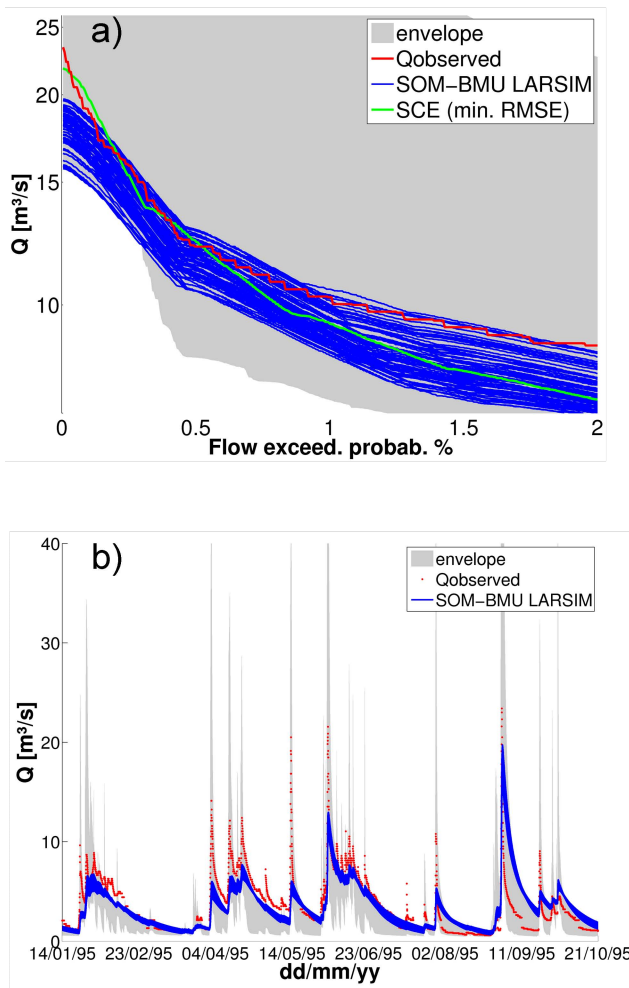
Fig. 13. LARSIM: **(a)** Flow duration curves (upper 2% section): Simulations corresponding to the BMU compared to the observed discharge and the results of an optimization approach using SCE-UA. **(b)** Time series corresponding to the BMU compared to the observed discharge.

Fig. 14. LARSIM: Results for the validation event August 2002 (BMU realizations and SCE-UA, RMSE).

hand corner of the map further corroborates this finding. The resulting model behaviour is illustrated in Fig. 13a and b. The envelope of the simulations in Fig. 13a shows that, in accordance with Fig. 11, a significant proportion of simulation runs (i.e. parameter combinations) is potentially capable of reaching sufficiently high and even excessive peak discharge values. However, the constraints linked to the lower (0.42–2%) part of the FDC counteract this behaviour with very high probability and force the position of the BMU (Fig. 11) towards the upper left hand side of the map. This finding is further supported by the results obtained by using the SCE-UA optimization algorithm (Duan et al., 1992) to minimize the RMSE of the simulated time series. In addition, the sharp gradients and extreme overestimation of *%DiffMaxPeak* and *%BiasFDC*$_{high}$ in reaction to changes of *EQD* and *EQD2* as well as the position of the simulation envelope in relation to
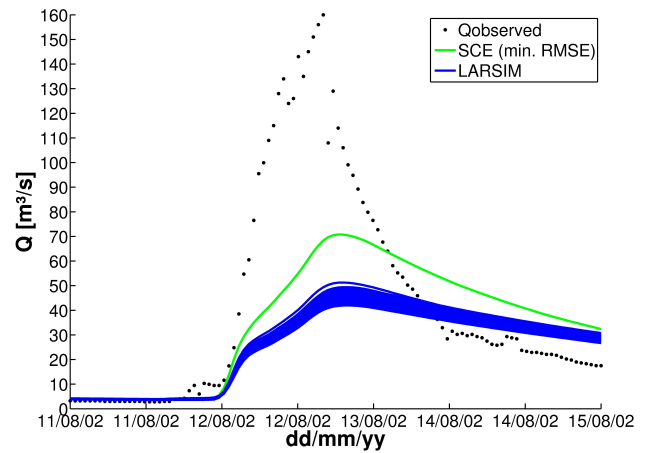
the observations (Fig. 13a) point at that the ranges allowed to these parameters in the course of the Monte-Carlo simulation are disproportionate. Consequently, the BMU parameter sets comprise high values for both *EQD* and *EQD2*, apparently in order to compensate for an excess in fast runoff components. This excess could have been triggered by of the settings for parameter *A*2 and/or overly high values for parameter *BSF*. The model realizations selected by using the BMU criterion display some kind of "plateau behaviour" which is illustrated by the decrease in slope towards the upper end of the FDC (Fig. 13a) and indicates deficits in discharge volume generation for runoff peaks. These deficits also manifest themselves with regard to the time series results for the training period (Fig. 13b) and, even more, for the extreme event of August 2002 (Fig. 14). Here, the model realizations that were attributed to the BMU perform even worse than the reference simulation which was obtained using the SCE-UA algorithm and hardly reach about 50 m^3/s peak flow compared to approximately 160 m^3/s of observed peak discharge.

3.1.3 WaSIM-ETH

Regarding the WaSIM-ETH model realizations, the indices used to examine the model behaviour show a marked correlation and a general gradient extending from the upper left to the lower right corner (Fig. 15). Nevertheless, the index ranges covered by the Monte-Carlo simulation with WaSIM-ETH are quite individual and comprise negative as well as positive values. Thus, the upper third of the map is generally characterized by underestimation of the indices which gradually fades to high index values towards the lower right hand side of the map such that the lowest index values (i.e. the model realizations that best "fit" the observations), and thus the BMU, can be found only a few nodes below the centre of the map.
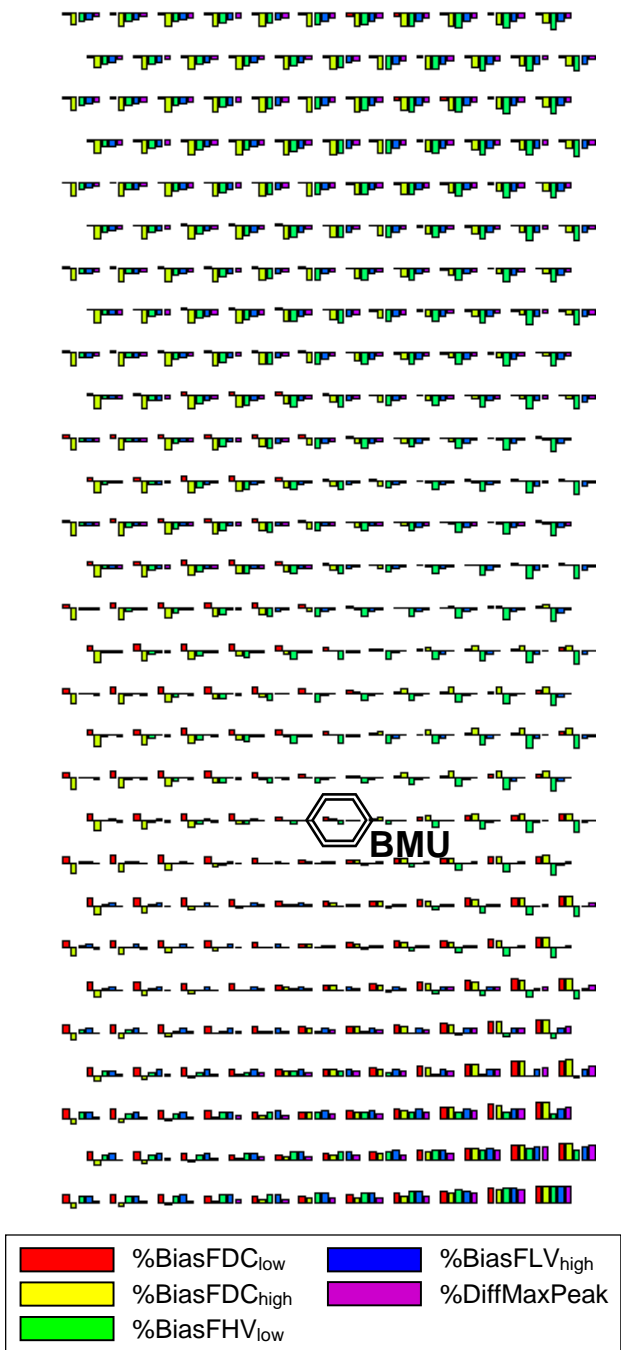
Fig. 15. WaSIM: Distribution of index properties on the map displayed as bar plots and the position of the best-matching unit (BMU). Note that the bars have individual relative scales.

Fig. 16. WaSIM: **(a)** Parameter plane, i.e. mean values of each model parameter for the simulations projected onto the individual map elements. **(b)** Distribution of reference vector (i.e. indices) properties on the map. The position of the BMU is marked.

Contrary to the other models we examined, from a visual comparison of Fig. 16a and the component planes (Fig. 16b) it is not possible to isolate any straightforward relationship between individual model parameters and indices, although each of the parameters we included in the Monte-Carlo simulation shows a marked sensitivity with respect to the index
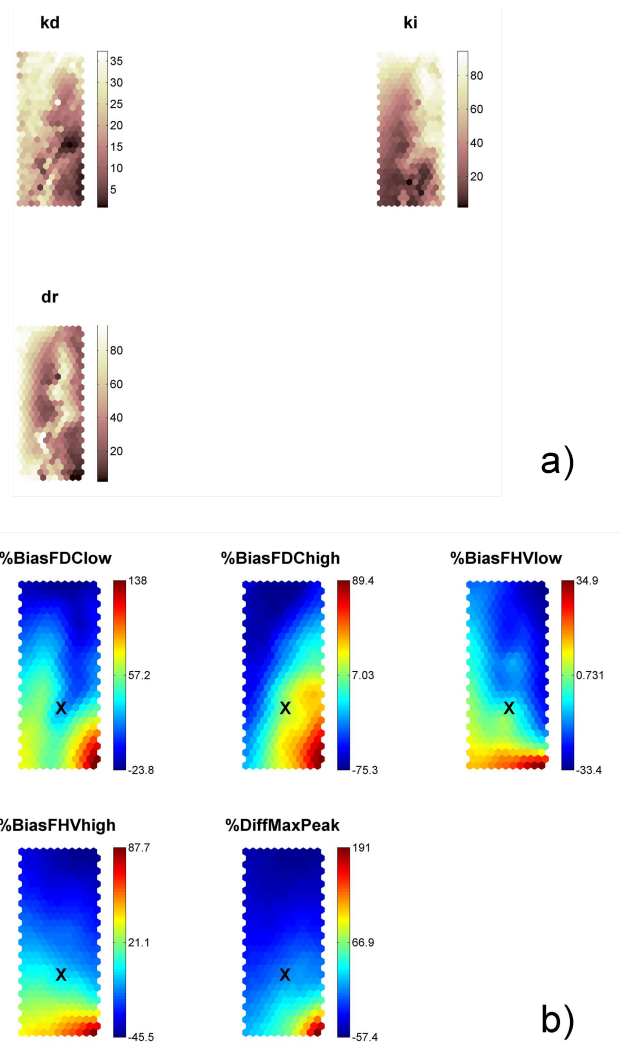
measures. However, to some extent, the parameter planes themselves (Fig. 16a) display correlated patterns, e.g. the map units with high values for *ki* suspiciously coincide with the map units for which parameter *kd* acquires predominantly low values. As to the general behaviour of the WaSIM-ETH model structure with respect to the indices, it can be stated that low values for parameter *kd* and *dr* in combination with intermediate *ki* values redound to an increase of peak flow.

From these findings we infer that all three model parameters are equally important for matching the five index measures and that parts of the model structure exhibit a strongly interacting, maybe even equifinal behaviour. That way, the effect of one parameter can be compensated to some extent by a combination of the other parameters. This would also provide some explanation for the fact that the BMU is located
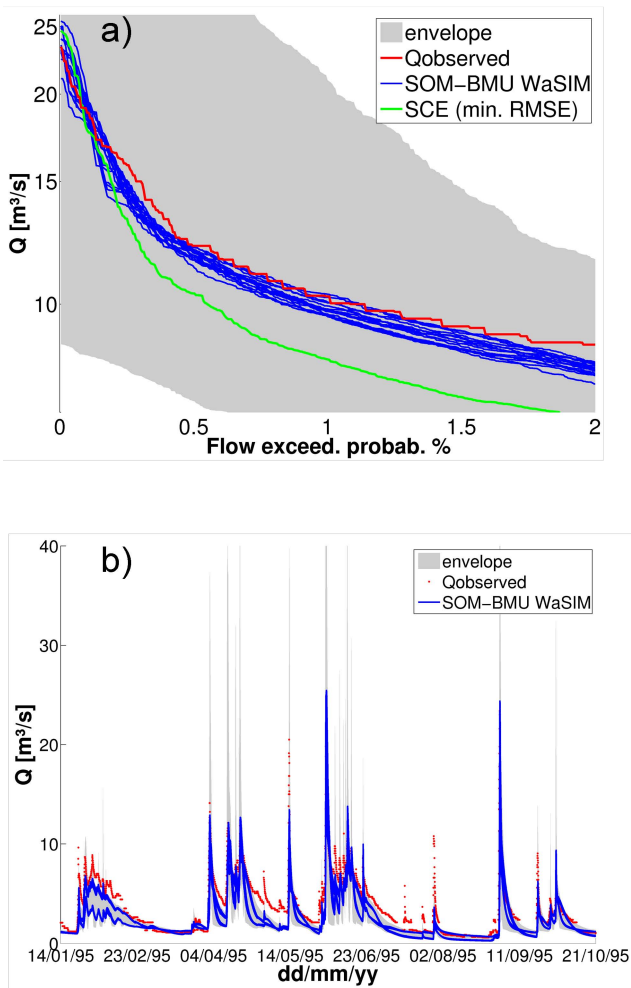
Fig. 17. WaSIM: **(a)** Flow duration curves (upper 2% section): Simulations corresponding to the BMU compared to the observed discharge and the results of an optimization approach using SCE-UA. **(b)** Time series corresponding to the BMU compared to the observed discharge.

Fig. 18. WaSIM: Results for the validation event August 2002 (BMU realizations and SCE-UA, RMSE).

tics, especially regarding the highest runoff values (Fig. 17a) whereas the FDC of the SCE-UA reference simulation adopts a steeper trajectory and hits the ordinate somewhat above the FDC of the observations. The model time series (Fig. 17b) finally exemplify that the BMU model realizations were obtained using parameter sets from disjoint regions of the parameter space. The peaks, however, are reproduced very well, with the exception of only a few minor runoff events. In contrast, the peak of the validation event (Fig. 18) is strongly underestimated by all model realizations, whereas only one half of the BMU model realizations exceed the peak flow of the SCE-UA reference simulation.

3.2 Results generated from SOM of the joint model data set

According to the initialization procedure (Sect. 2.4) the number of neurons on a map does not increase linearly with the number of data items used for the training. Therefore, the proportion of data items to the number of map units results somewhat higher for the SOM trained on the joint data set than for the SOMs we discussed in the previous section. The overall quantization error $\bar{d}=0.38$ (Eq. 10) of this SOM result appears to be sufficiently low so as to characterize the mapping as a very good approximation of the model data.

In Fig. 19a the neurons of the map are reproduced as pie charts in order to illustrate the distribution of data items from the different models on the SOM. These represent the percentage of data from each model that has been attributed to the neurons via the training. In Fig. 19b the same pie charts are scaled using the number of data items on each map unit in order to simultaneously visualize the distribution of data quality and quantity on the map. The void regions on the map indicate interpolative units where the data items are clearly disjoint and characterized by marked differences with regard

in a map region where all three parameters are subject to considerable alterations. It finally results that the parameter sets associated to the BMU can be divided into two groups with strongly contrasting values: The first group has low values with respect to kd and simultaneously high values for dr while the second group displays high kd values in combination with low values for dr. The respective index measures, however, turned out to be quite similar.

With $\bar{d}_{\text{BMU}}=0.76$ the quantization error of the BMU model realizations that correspond to the time series of measured discharges is the smallest among the three models and indicates that these parameter sets match the observed behaviour of the system relatively well. This is further supported by the rather central position of the BMU on the map. The model realizations which are retrieved from the BMU thus result in a rather accurate representation of the FDC characteris-
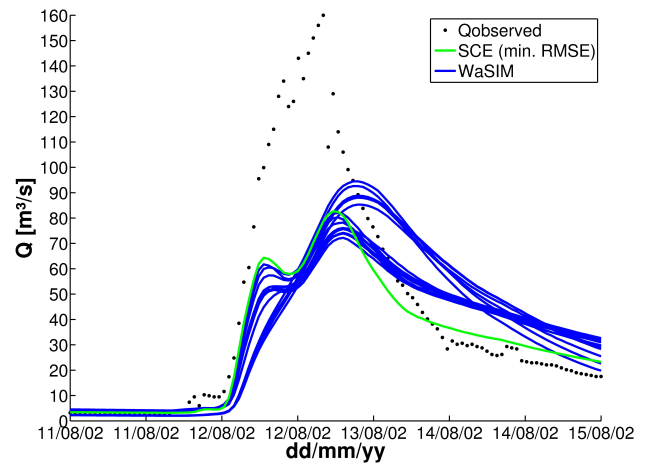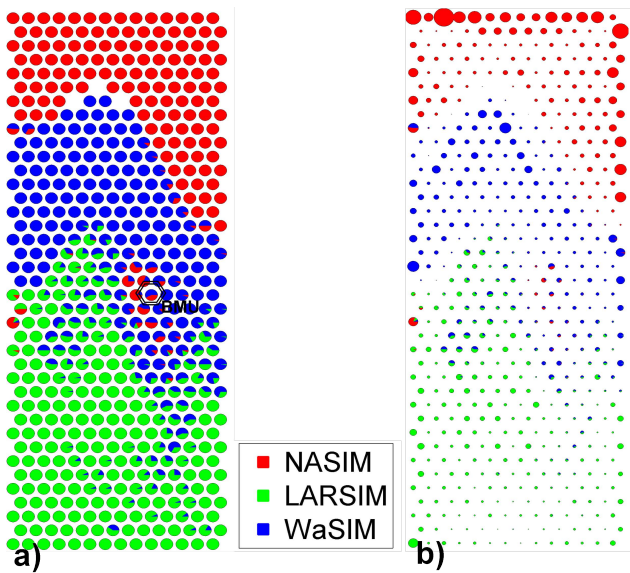
Fig. 19. Comaparison of NASIM, LARSIM and WaSIM: **(a)** The neurons of the SOM are reproduced as pie charts that represent the percentage of data from each model that has been attributed to the neurons via the training. **(b)** same as (a) but displayed as a "hit histogram", i.e. the size of the pie charts is proportional to the number of data items attributed to the corresponding neuron.

Fig. 20. SOM of the combined data from all models in bar plot illustration.

to their indices. With the exception of some very isolated occurrences there are no nodes on the map that are simultaneously populated with model realizations from all three models. The same holds true for simultaneous occurrences of NASIM and LARSIM realizations on a node. Close to the center of the map the nodes are predominantly populated with mixtures of model realizations from LARSIM and WaSIM-ETH as well as WaSIM-ETH and NASIM. Following the theory of Self-Organizing Maps (Sects. 2.3 and 2.5), it can be assumed that these model realizations display equivalent characteristics with respect to the indices we used to describe them. The distances between the nodes further allow inferring that the differences between NASIM and LARSIM in terms of index characteristics are stronger than the differences between realizations of NASIM and WaSIM-ETH or LARSIM and WaSIM-ETH, which is most importantly highlighted by the fact that hardly any node is populated at the same time with realizations from NASIM and LARSIM. The neurons close to the bottom of the map, on which simultaneous occurrences of LARSIM and WaSIM-ETH can be found, point at sporadic extremes of WaSIM-ETH model behaviour. Figure 19b further exemplifies that similarities between model realizations from the three models only occur with rather low probability. Moreover, it can be seen that the data items are distributed with a higher density around the upper margin of the map. There are two potential possibilities that lend themselves to explain this phenomenon: When the multidimensional data distribution is rapidly thinning to-
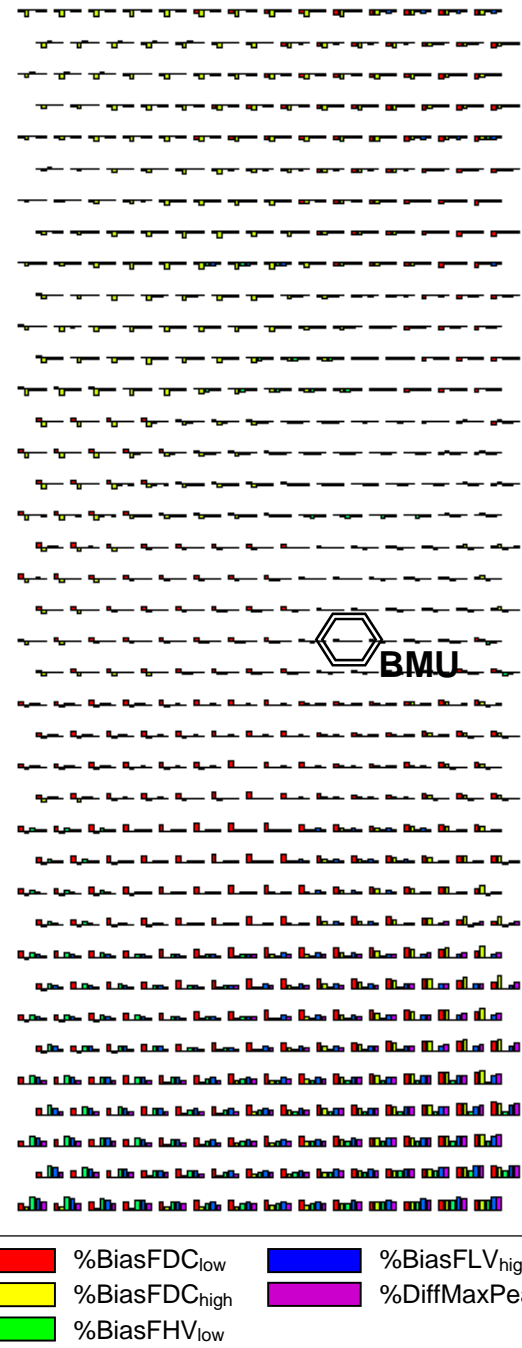
wards its margins it can not always be covered entirely by the SOM reference vectors. Consequently, the remaining data sets are attributed to the nearest, marginal nodes. Otherwise, this type of distribution could indicate that the corresponding model realizations are indeed very similar compared to the remaining data set and thus attributed to the same neuron.

The aforementioned individuality of model realizations is further demonstrated by the bar plot of the map (Fig. 20)

which compares the index characteristics that distinguishes the models. The mapping shows a very distinct organization of data properties: Predominantly underestimated index values on the upper part gradually fade to overestimated indices in the lower part. The model characteristics that are captured in this representation of the map correspond exactly with Figs. 7, 11 and 15. The position of the BMU which represents the most likely location of the measured time series on the map, is already roughly identifiable from the bar plots in Fig. 20 and coincides, according to Fig. 19a, with the map region in which the neurons are simultaneously populated with data from the models NASIM and WaSIM-ETH. Thus, these models can be characterized as equivalent, of course, only according to the criteria which have been imposed in our study to discriminate between individual model realizations.

The model realizations allocated to the BMU partly correspond with the BMU realizations which have been identified using the mappings of the individual data sets. However, as a consequence of the map dimensions, a somewhat higher number of model realizations are attributed to this BMU. Nevertheless, it can be seen from the FDC plot in Fig. 21a and from Fig. 21b and c that these model realizations still represent the characteristics of the observed discharge time series very well. A distinctive feature of the model realizations obtained from NASIM (Fig. 21b) is the partial overestimation of peak discharge which is why a better performance of these realizations with respect to the extreme flood event from August 2008 can be expected, compared to the results from Sect. 3.1. This effect, however, can not be influenced deliberately and must be attributed to the wider range of data items on the BMU.

4 Discussion and conclusion

Similar to Herbst et al. (2009) our study is based on a combined approach: While the indices adopt the function of performance measures (rather than Signature Measures, according to their underlying theory, see Gupta et al., 2008) the Self-Organizing Map serves as a tool to analyze and visualize the data which is obtained through them.

The indices we used were conceptualized to extract data on very specific characteristics of model behaviour according to the focus of our study. These characteristics are represented by a choice of FDC-based indices that are intended to focus on the reproduction of peak discharges. They consequently have to be understood as an example of a model evaluation problem. Of course, the choice of indices can be tailored, according to the individual goals of the model analysis. This also includes a weighting of individual indices or measures in order to put more emphasis on specific time series features. However, we did not make use of this option and preferred to weight all indices with 1 instead. Thus, our study shares the underlying assumptions of the work by
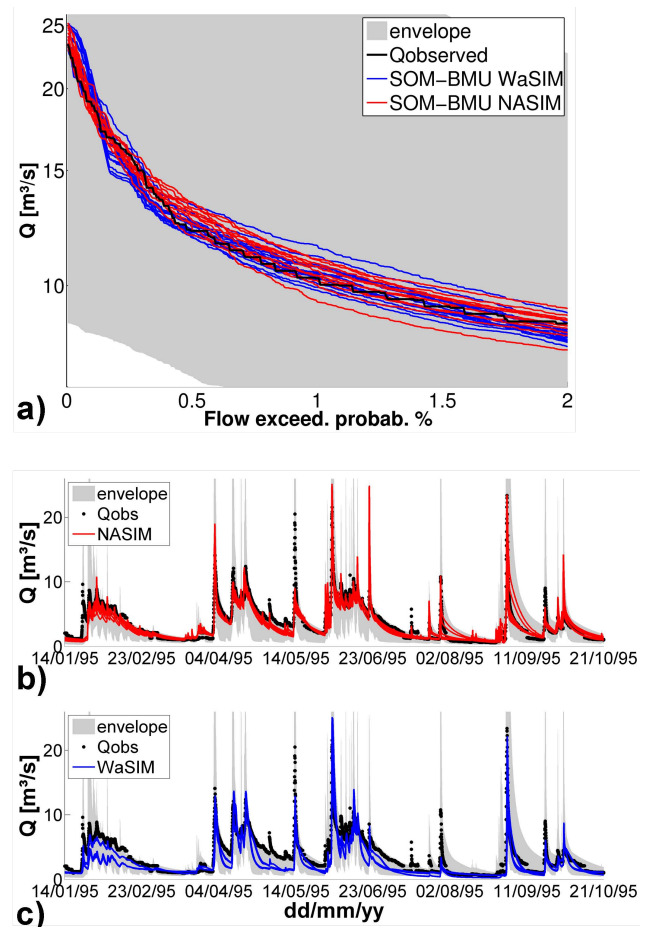

Fig. 21. (a) Flow duration curves (upper 2% section): Simulations corresponding to the BMU of the SOM trained on the combined model data compared to the observed discharge and the results of an optimization approach using SCE-UA. **(b)** Time series of the model NASIM corresponding to the BMU of the SOM trained on the combined model data compared to the observed discharge. **(c)** Time series of the model WaSiM-ETH corresponding to the BMU of the SOM trained on the combined model data compared to the observed discharge.

Herbst et al. (2009), namely that the index or performance measures are equally relevant and that the model is capable of reproducing them.

The SOM helps to analyze the data obtained via the indices by producing a discretized (and thus data-compressed) mapping of their distribution in the index space onto a two dimensional plane such that their pattern and consequently the patterns of model behaviour can be conveyed in a comprehensive manner. This is achieved by different visualization techniques (see also Vesanto, 1999) and importantly by linking the model properties to the corresponding parameter space. In a sense, the SOM helps to turn the data extracted via the indices into information on model behaviour which subsequently can be used in the decision making process.

The results from Sects. 3.1 and 3.2 clearly demonstrate that a SOM can be used to cluster model output data according to different (time series) characteristics. Although the indices used in our study are not fully independent (finally they are all derived from the FDC) they effectively helped differentiating the simulation results obtained from the watershed models NASIM, LARSIM and WaSIM-ETH. It has been demonstrated that the clustering of model output data provides useful insights, such as a preliminary sensitivity analysis and a general characterization of model behaviour regarding the reproduction of peak discharges. In addition, the presented approach allows identifying the model parameters and -time series that best approximate the observations with respect to a given set of constraints embodied by the indices. This is achieved by determining the BMU of the index vector that corresponds to the observations, which involves identifying the reference vector that minimizes the Euclidean distance to the observations vector. This procedure, commonly used in many SOM applications, deserves critical attention as it implies converting a multi-objective optimization to a single-objective problem (e.g. Madsen 2003) which does not always permit to find the optimal solution of a multi-objective problem (Zadeh, 1963; see also Gupta et al., 1998).

Notwithstanding, the results obtained for the BMU of the NASIM map and the WaSIM-ETH map represent the characteristics of the observed time series with similar or partially superior accuracy compared to the results we obtained by implementing a simple calibration strategy by means of the optimization algorithm SCE-UA. This finding, on one hand, is partly owed to the fact that the SCE-UA optimization algorithm, in contrast to the SOM approach, allows to globally searching the parameter space with potentially infinite resolution. On the other hand, the poor results with respect to the BMU of the LARSIM map are attributable to the constraints that were imposed by applying the five index measures. These constraints turned out to be overly rigorous to be simultaneously satisfied by the LARSIM model and are probably incompatible with its general behaviour. Consequently, the given set of indices avoided that its potential could be exploited to the full extent. At this point it has to be stressed that an accurate reproduction of runoff events in the first place depends on the quality of the precipitation input data. However, as to this aspect, the results obtained with the three models give no reason for concern. Besides the information on model behaviour that can be extracted using the SOM approach, one of its strengths has to be seen especially in the ability to extract a set of model parameters that meet a set of very specific criteria (which e.g. could have been imposed by decision makers). The corresponding parameter ranges, in turn, constitute a potential key for the assessment of parameter uncertainties.

Regarding the simulation of the extreme flood event (11–15 August 2008) our approach did not yield clear improvements compared to the much simpler SCE-UA calibration strategy. Thus, it has to be put into question whether the predictive abilities of hydrological models can be enhanced using this approach. In contrast, the results rather indicate that the ability of a model to "extrapolate" to behavioural domains beyond the calibration data can be exploited with a higher probability if the model realizations match or overestimate the highest section of the FDC. Another, rather self-evident conclusion from the LARSIM result is that a successful calibration strategy always has to consider the peculiarities of model behaviour. However, this behaviour is also largely determined by the parameters which, in the scope of our study, are considered as constant, among other important influences such as the input data. Thus, the parameterization used for the LARSIM model deserves further critical analysis in order to elucidate its behaviour.

The most prominent advantage of the SOM in the context of model analysis is that it allows to simultaneously evaluating the data from two or more models. Using a SOM in combination with an appropriate set of "measures" that help extracting specific information from time series, model realizations that satisfy a given set of criteria can be selected from among various model structures at a time. As only similar data items are attributed to the same map unit (see Sects. 2.3 and 2.5) the distribution in Fig. 19a, on one hand, highlights the individuality of the different model data sets with regard to their behaviour which is expressed through characteristic combinations of index values. On the other hand, it provides a vivid evidence of the high discriminatory power of the SOM approach.

The possibility of a direct comparison of model behaviour properties lends itself for a series of potential applications, e.g. in a model ensemble framework: The proportion of "equivalent" model realizations in a data set obtained from the results of an ensemble simulation, in turn, could serve as a "proxy" for the independence of model structures. That way, a set of model realizations or model structures that together cover a broader range of measured system behaviour than each individual model (e.g. model realizations that emphasize different sections of the FDC) could be determined and constitute the base of a (multi-)model ensemble application (Fenicia et al., 2007).

Edited by: A. Schumann
Reviewed by: two anonymous referees

References

Bossel, H.: Systeme, Dynamik, Simulation – Modellbildung, Analyse und Simulation komplexer Systeme, Books on Demand, Norderstedt, 2004.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, Water Resour. Res., 36, 3663–3674, 2000.

Bremicker, M.: Das Wasserhaushaltsmodell LARSIM. Modellgrundlagen und Anwendungsbeispiele, Freiburger Schriften zur Hydrologie, 2000.

Casper, M. C., Herbst, M., Grundmann, J., Buchholz, O., and Bliefernicht, J.: Einfluss der Niederschlagsvariabilität auf die Simulation extremer Abflüsse, Hydrology and Water Resources Management Germany, accepted, 2009.

Disse, M., Pakosch, S., and Yörük, A.: Development of an operational expert system for flood forecasts considering prediction uncertainty, Hydrologie und Wasserbewirtschaftung, 51, 210–215, 2007.

Duan, Q.: Global optimization for watershed model calibration, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application, AGU, Washington D.C., 89–104, 2003.

Duan, Q., Sorooshian, S., and Gupta, V. K.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28, 1015–1031, doi:10.1029/91WR02985, 1992.

Fenicia, F., Solomatine, D. P., Savenije, H. H. G., and Matgen, P.: Soft combination of local models in a multi-objective framework, Hydrol. Earth Syst. Sci., 11, 1797–1809, 2007, http://www.hydrol-earth-syst-sci.net/11/1797/2007/.

Gan, T. Y. and Biftu, G. F.: Effects of model complexity and structure, parameter interactions and data on watershed modelling, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., Water Science and Application, AGU, Washington D.C., 317–329, 2003.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour. Res., 34, 751–764, 1998.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22, 3802–3813, doi:10.1002/hyp.6989, 2008.

Haykin, S.: Neural networks – a comprehensive foundation, 2nd ed., New Jersey, 842 pp., 1999.

Herbst, M. and Casper, M. C.: Towards model evaluation and identification using Self-Organizing Maps, Hydrol. Earth Syst. Sci., 12, 657–667, 2008, http://www.hydrol-earth-syst-sci.net/12/657/2008/.

Herbst, M., Gupta, H. V., and Casper, M. C.: Mapping model behaviour using Self-Organizing Maps, Hydrol. Earth Syst. Sci., 13, 395–409, 2009, http://www.hydrol-earth-syst-sci.net/13/395/2009/.

Hydrotec: Rainfall-Runoff-Model NASIM – program documentation (in German), Hydrotec GmbH, Aachen, 579 pp., 2005.

Kalteh, A. M., Hjorth, P., and Berndtsson, R.: Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application, Environmental Modelling & Software, 23, 835–845, doi:10.1016/j.envsoft.2007.10.001, 2008.

Kaski, S.: Data Exploration Using Self Organizing Maps, Dr. thesis, Department of Computer Science and Engineering, Helsinki University of Technology, Helsinki, 57 pp., 1997.

Klemeš, V.: Conceptualization and scale in hydrology, Journal of Hydrology, 65, 1–23, 1983.

Kohonen, T.: Self-Organizing Maps, 3rd ed., Information Sciences, Springer, Berlin, Heidelberg, New York, 501 pp., 2001.

Legates, D. R. and McCabe Jr., G. J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35, 233–241, 1999.

Kundzewicz, Z. W., Mata, L. J., Arnell, N. W., Döll, P., Kabat, P., Jiménez, B., Miller, K. A., Oki, T., Sen, Z., and Shiklomanov, I. A.: Freshwater resources and their management, in: Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J., and Hanson, C. E., Cambridge University Press, Cambridge, 173–210, 2007.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Adv. Water Resour., 26, 205–216, 2003.

Merz, B. and Didszun, J.: Risikomanagement extremer Hochwasserereignisse, USWF – Zeitschrift für Umweltchemie und Ökotoxikologie, 17, 191–192, 2005.

Merz, B., Didszun, J., and Ziemke, B.: RIMAX Risikomanagement extremer Hochwasserereignisse, 2. Auflage ed., GeoForschungsZentrum Potsdam (GFZ), Potsdam, 51, 51 pp., 2007.

Minns, A. W. and Hall, M. J.: Artificial Neural Network Concepts in Hydrology, in Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G., 307–320, 2005.

Moore, C. and Doherty, J.: Role of the calibration process in reducing model predictive error, Water Resour. Res., 41, W05020, doi:10.1029/2004WR003501, 2005.

Peschke, G.: Soil Moisture and Runoff Components from a Physically Founded Approach, Acta hydrophysica, 31(3/4), 191–205, 1987.

Reusser, D. E., Blume, T., Schaefli, B., and Zehe, E.: Analysing the temporal dynamics of model performance for hydrological models, Hydrol. Earth Syst. Sci. Discuss., 5, 3169–3211, 2008, http://www.hydrol-earth-syst-sci-discuss.net/5/3169/2008/.

Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, in: Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G., Wiley, 193–220, 2005.

Schulla, J. and Jasper, K.: Model description WaSIM-ETH, Zürich, 167 pp., 2001.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., and Gupta, H. V.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis, Hydrol. Process., 17, 455–476, 2003.

Van Genuchten, M. T.: A Closed-Form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils, Soil Sci. Soc. Am. J., 44(5), 892–898, 1976.

Vesanto, J.: SOM-based data visualization methods, Intelligent Data Analysis, 3(2), 111–126, 1999.

Vesanto, J.: Neural network tool for data mining: SOM toolbox, Symposium on Tool Environments and Development Methods

for Intelligent Systems (TOOLMET2000), Oulu, Finnland, 184–196, 2000.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J.: SOM Toolbox for Matlab 5, Helsinki University of Technology, Report A57, Espoo, Finland, 60 pp., 2000.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res., 44, W09417, doi:10.1029/2007WR006716, 2008.

Zadeh, L.: Optimality and non-scalar-valued performance criteria, IEEE T. Automat. Contr., 8, 59–60, 1963.

Zhao, R. J.: Flood forecasting method for humid regions of China, East China College of Hydraulic Engineering, Nanjing, 1977.

ANGABEN ZUR PERSON

Vorname, Name	Marcus Herbst
Geburtsdatum	29.12.1974 (Cuxhaven)
Familienstand	ledig

BERUFSERFAHRUNG

seit Mai 2004	Wissenschaftlicher Mitarbeiter an der Universität Trier im Fach Physische Geographie, Arbeitsgruppe „Modellbildung und Simulation von Geoprozessen"

AUSBILDUNGSWEG

Januar 2006 – Juli 2009	Doktorand an der Universität Trier, Fachbereich VI – Geographie/Geowissenschaften
April 1997 – September 2003	Studium der Geographie an der Christian-Albrechts-Universität zu Kiel, Abschluss: Diplom-Geograph Nebenfächer: Landschaftsentwicklung, Geologie Titel der Diplomarbeit: „Integration von Methodenbausteinen in ein GIS-gestütztes Landmanagementsystem: Einsatz von ArcGIS für die automatisierte Abschätzung der Landschaftsfunktionen Grundwasserneubildung, Nitratrückhaltevermögen und Schwermetallfilterfunktion" Betreuer: Prof. Dr. R. Duttmann
Oktober 1995 – März 1997	Studium der Physik an der Christian-Albrechts-Universität zu Kiel (abgebrochen)
1988 – 1995	Lichtenberg-Gymnasium Cuxhaven, Abschluss: Abitur