

Towards model evaluation and identification using Self-Organizing Maps

M. Herbst and M. C. Casper

Department of Physical Geography, University of Trier, Germany

Received: 17 October 2007 – Published in Hydrol. Earth Syst. Sci. Discuss.: 5 November 2007

Revised: 13 February 2008 – Accepted: 17 March 2008 – Published: 9 April 2008

Abstract. The reduction of information contained in model time series through the use of aggregating statistical performance measures is very high compared to the amount of information that one would like to draw from it for model identification and calibration purposes. It has been readily shown that this loss imposes important limitations on model identification 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 discriminatory 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 analyzed and compared with the manually calibrated model as well as with the results of the Shuffled Complex Evolution 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 therefore 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 measures 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 certain 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 aggregated into a single numerical value, regardless of the characteristic 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 parameter sets used to generate them are widely scattered throughout the parameter space. Therefore the information conveyed by aggregating goodness-of-fit measures is likely to be insufficient 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 complementary 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 evaluation 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 identifiability. 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 $\mathbf{x} \in X$ is considered as a vector

$$\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathfrak{R}^n \quad (1)$$

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

$$\mathbf{m}_i = [\mu_{i1}, \mu_{i2}, \dots, \mu_{in}]^T \in \mathfrak{R}^n \quad (2)$$

also called reference vector, which has the same dimensionality as the input vectors $\mathbf{x} \in X$. These weights connect each input vector \mathbf{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 \mathbf{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 $\mathbf{x} \in X$ and compute the Euclidean distance

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

between \mathbf{x} and each of the reference vectors \mathbf{m}_i (as a measure of similarity, generally any other metric can be applied as well) and find the neuron $c(\mathbf{x})$ with a reference vector \mathbf{m}_c such that

$$\|\mathbf{x} - \mathbf{m}_c\| = \min_i \{\|\mathbf{x} - \mathbf{m}_i\|\} \quad (4)$$

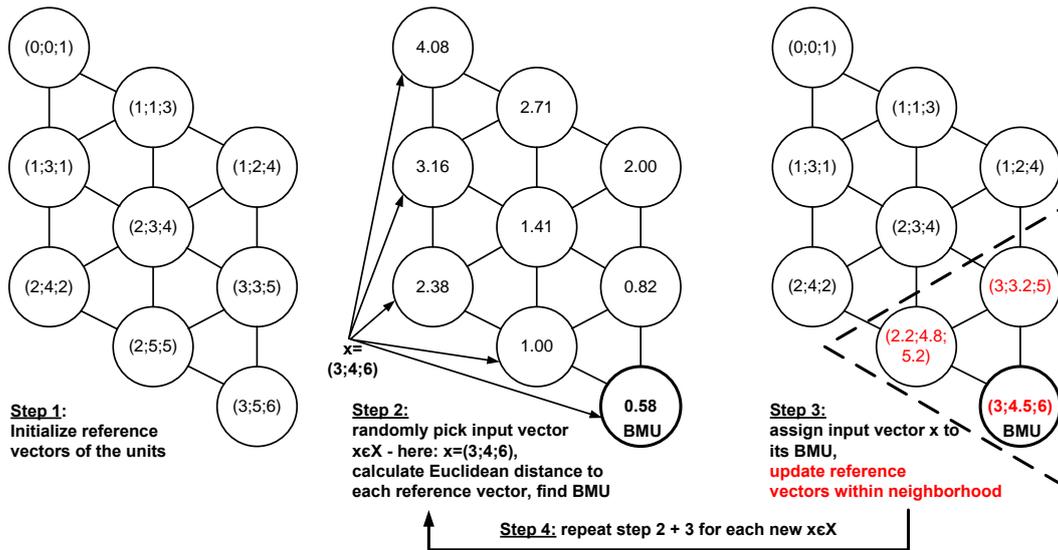


Fig. 1. The basic steps of the SOM algorithm.

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

$$m_i(t + 1) = m_i(t) + \alpha(t) h_{ci}(t) [x(t) - m_i(t)] \quad (5)$$

where t is the number of the iteration step and $m_i(t)$ is the current weight vector which is updated proportionally to the difference $[x(t) - m_i(t)]$. $h_{ci}(t)$ determines the degree of neighbourhood between the winning neuron c and neuron i for an input $x \in X$, 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

$$h_{ci}(t) = \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \quad (6)$$

is used, whereas $\|r_c - r_i\|^2$ 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 $h_{ci}(t) \rightarrow 0$ for $t \rightarrow \infty$. 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 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 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 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 m_i can be interpreted as pointers for their corresponding neuron into the input space X , hence the elements of 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

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

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

vector \mathbf{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 $\mathbf{m}_{c(y)}$ is activated for which

$$\|\mathbf{y} - \mathbf{m}_{c(y)}\| = \min_i \{\|\mathbf{y} - \mathbf{m}_i\|\} \quad (7)$$

The “image” $c(y)$ of the projected data item \mathbf{y} then represents the domain of input data patterns from X that is most similar to \mathbf{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 \mathbf{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

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

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

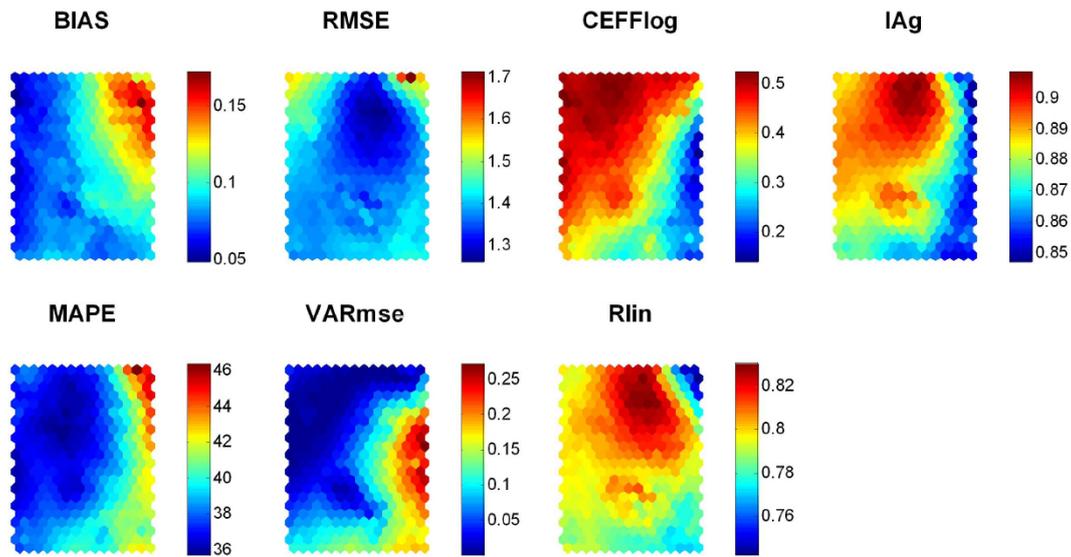


Fig. 2. Distribution of the mean values of each performance measure from Table 2 over the SOM lattice.

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)

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

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 km² low-mountain range test watershed “Schwarze Pockau” Saxony (Germany), situated near the border to Czechoslovakia and tributary of the Freiburger 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.

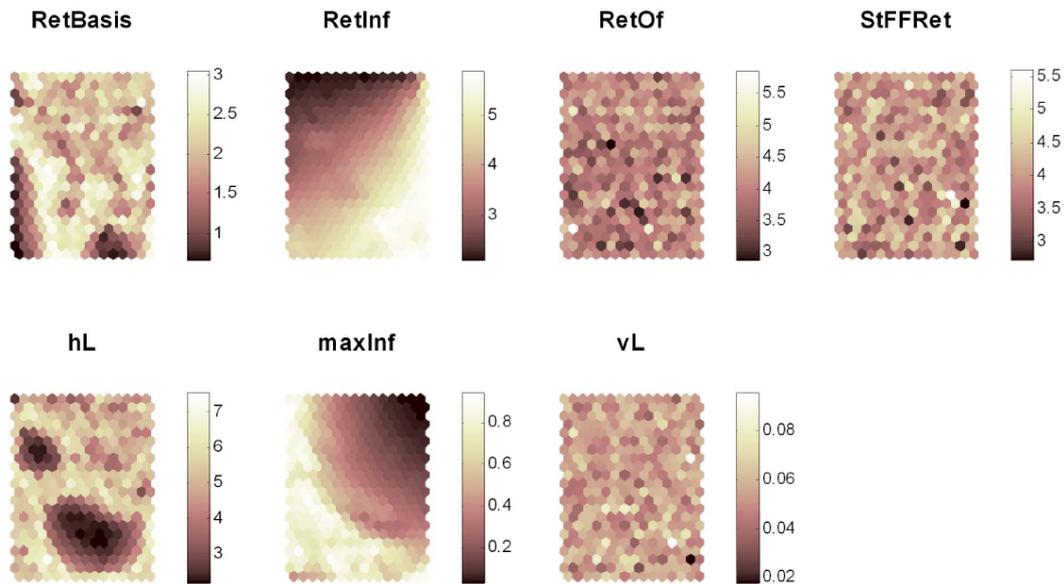


Fig. 3. Distribution of the mean values of each model parameter from Table 1 over the SOM lattice.

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

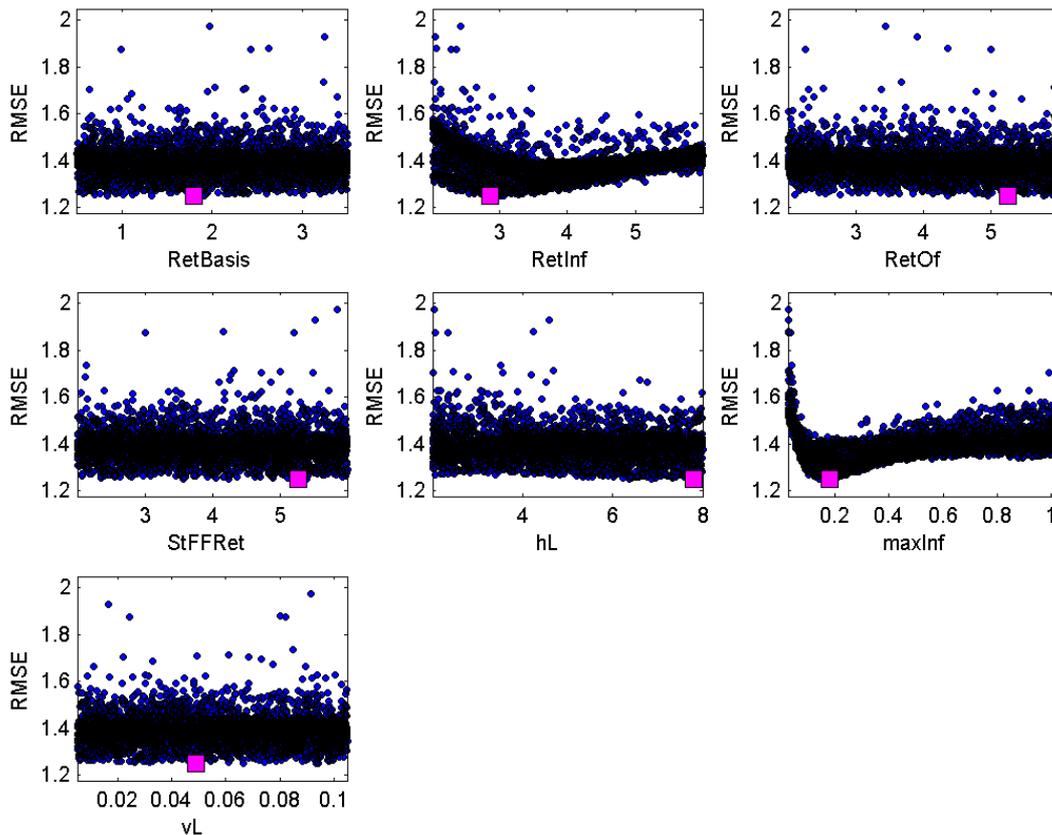


Fig. 4. Scatterplots of RMSE values for each of the examined NASIM parameters.

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

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

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

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,

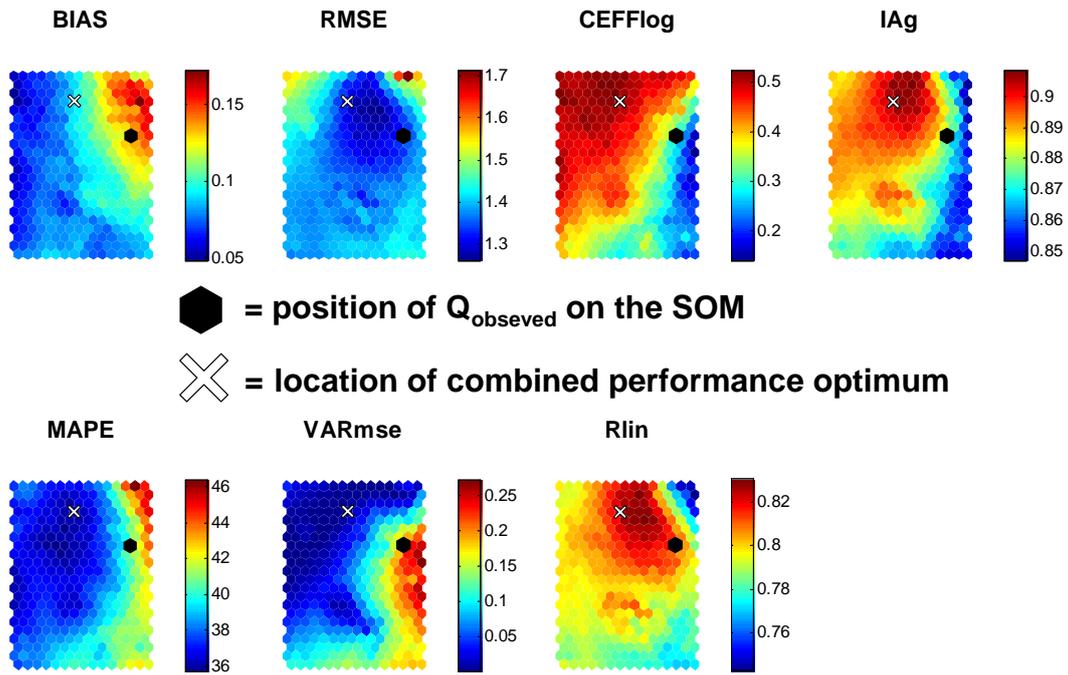


Fig. 5. The location of the best-matching unit (indicated by the black dot) for an input vector that represents the measured discharge time series. The white cross marks the common optimum (balance point) of the seven performance measures on the SOM grid.

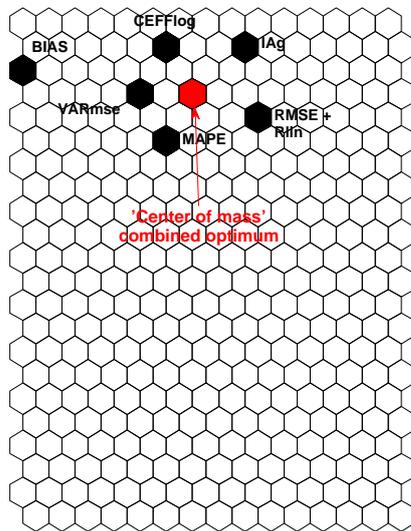


Fig. 6. Determination of the common optimum location (balance point) as the geometric center of mass resulting from the locations of the seven objective function optima on the SOM. Each optimum has unit weight. Here the RMSE and Rlin optima coincide on the same position.

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 max-Inf, 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

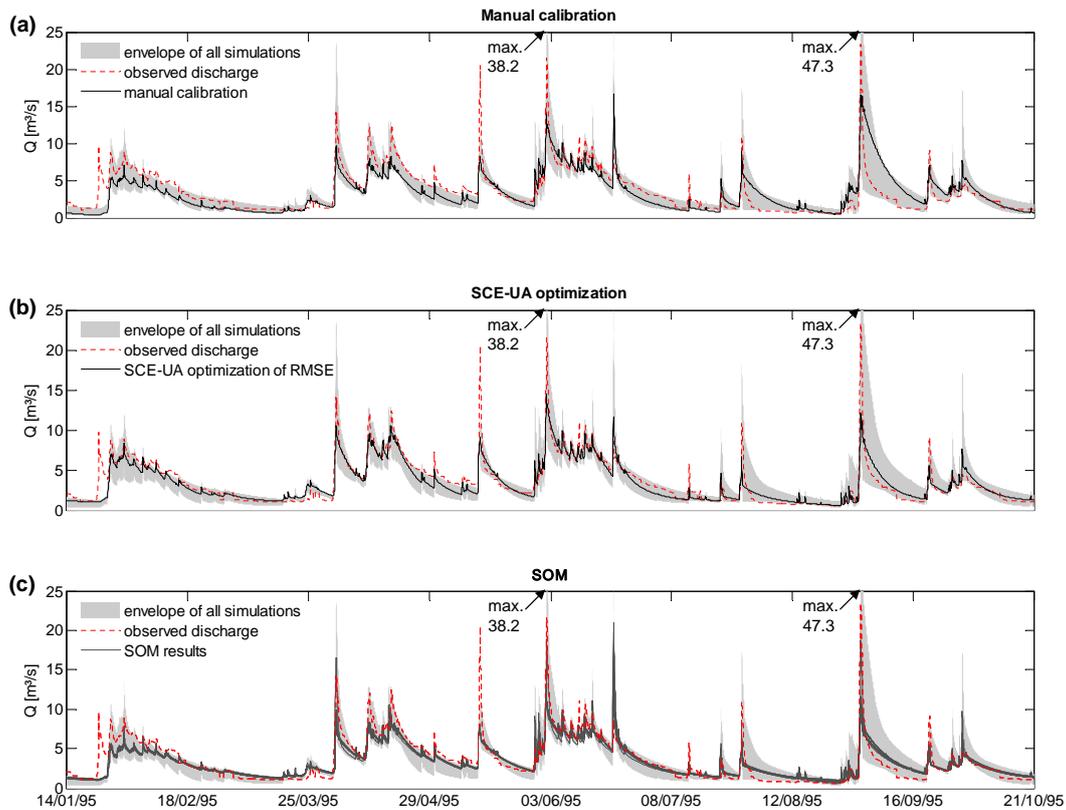


Fig. 7. The model realizations as resulting from (a) manual calibration, (b) optimization with the SCE-UA algorithm and (c) the BMU of the SOM for the measured discharge time series. The time-series are compared to the measured discharge and the envelope of the Monte-Carlo simulation, i.e. the area which 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.

Acknowledgements. The authors wish to thank O. Buchholz (Hydrotec GmbH) and R. Wengel for their support. This work was carried out using the SOM-Toolbox for Matlab by the “SOM Toolbox Team”, Helsinki University of Technology (<http://www.cis.hut.fi/projects/somtoolbox>).

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.-I., 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.

- 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.
- Wagner, T., Wheeler, 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.
- Wagner, T., Wheeler, 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.