iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Towards model evaluation using Self-Organizing Maps

M. Herbst and M. C. Casper

Department of Physical Geography, University of Trier. FBVI, Campus II, Behringstrasse, D-54286 Trier, Germany (herbstm/casper@uni-trier.de)

Abstract: Basically, any statement on hydrological model behaviour depends on our possibilities to differentiate between model time series. Applied within a model identification context, aggregating statistical performance measures are inadequate to capture details on time series characteristics as essentially different model results can be produced with close to identical performance measure values. It has been readily shown that the loss of information on the residuals 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 tentatively used as an alternative to a classical optimization algorithm to identify the model realizations among the Monte-Carlo simulations 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).

Keywords: SOM; Self-Organizing Map; model evaluation; optimization.

1. INTRODUCTION

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 however these measures involve considerable problems [Yapo et al., 1998; Lane, 2007]. 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. Because of their low discriminatory power traditional performance measures are not well suitable to give evidence of the difference or equivalence (i.e. equifinality) between alternative model realizations [Gupta et al., 2003; Beven and Binley, 1992]. This, in turn, implies serious limitations to model calibration and identification; in a sense it constitutes another source of model uncertainty [Wagener et al., 2003].
As a step toward improved extraction of information from existing data we introduce an approach that circumvents 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 data mining and process monitoring [Alhoniemi et al., 1999; Simula et al., 1999; Vesanto, 2000]. In the context of the hydrological sciences, however, applications of SOM are still rather uncommon.

2. METHODS

2. 1 Self-Organizing Map

A 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: It maps vectorial input data items with similar patterns onto contiguous locations of a discrete low-dimensional grid of neurons, i.e. it has no output function like other types of ANN. Nearby locations on this map are attributed similar data patterns. Thus, in the course of the training, each of the map's neurons is 'tuned' to a different domain of the patterns contained in the vectorial training data items. The map units act as decoder for different types of patterns contained in the input data [Kohonen, 2001]. An input data item $\boldsymbol{x} \in X$ is considered as a vector

$$\mathbf{x} = \left[x_1, x_2, \ldots, x_n\right]^T \in \Re^n \qquad (1)$$

with n being the dimension of the input data space. In our case n is the length of the time series. A fixed number of neurons is arranged on a regular grid whose dimensions can be determined by means of heuristic algorithms if no other preferences are made. Each neuron is being associated to a weight vector

$$\mathbf{m}_i = \left[\mu_{i1}, \mu_{i2}, \ldots, \mu_{in}\right]^T \in \Re^n \qquad (2)$$

also called reference vector, which has the same dimensionality as the input vectors $\boldsymbol{x} \in X$. These weights connect each input vector $\boldsymbol{x}$ in parallel to all neurons (indexed i) of the map. Moreover the neurons are connected to each other. In our case this interconnection is defined on a hexagonal grid topology. Fig. 1 explains the functioning of a SOM: In each iteration step the Euclidean distance between a randomly chosen input data item $\boldsymbol{x}$ and the reference vectors $\boldsymbol{m}_i$ is calculated. The neuron with minimal Euclidean distance to this data item is called the best-matching unit (BMU). Subsequently the reference vectors in the neighbourhood of this BMU are updated. However, the rate of change of the reference vectors decrease proportionally to the difference between $\boldsymbol{x}$ and $\boldsymbol{m}_i$ and the number of iteration steps. Moreover, also the radius of the neighbourhood decreases proportionally to the number iteration steps (commonly a Gaussian function is used to define the neighbourhood). The mapping "self-organizes" upon repeated cycling through the input data sets. For more detailed information on SOM and its properties please refer to Kohonen [2001] or Haykin [1999]. A concise description of the algorithm is given in Herbst and Casper [2008]. In this contribution we especially make use of the fact that SOM can also be applied to project an input data vector $\boldsymbol{y}$ onto the map which has not been part of the training data manifold. This means that the node with the reference vector $\boldsymbol{m}_c$ is selected for which the Euclidean distance between $\boldsymbol{y}$ and $\boldsymbol{m}_c$ is minimal. This "image" of the projected data item then represents the domain of input data patterns from X that is most similar to $\boldsymbol{y}$. Moreover, as the number of available neurons is much smaller than the number of vectors used for the training, the selected neuron will be associated to a range of input data patterns from X which will represent the domain of input data patterns that is closest to $\boldsymbol{y}$.

2.2 Data, preparation and experimental setup

In the present example 4000 model time series constituted the input data vectors of the training data set. These were obtained from Monte Carlo simulations with the distributed conceptual watershed model NASIM [Hydrotec, 2005] running at hourly time steps over a period of two years (1 November 1994 to 28 October 1996), i.e. each input data vector

consisted of 17472 elements. 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 as a black-box.

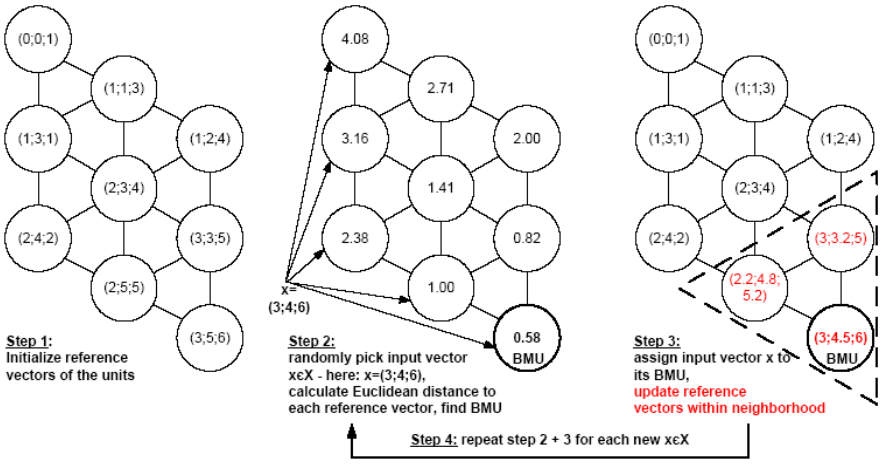

Figure 1. The basic steps of the SOM algorithm.

Seven model parameters were selected for Monte-Carlo random sampling (Tab. 2). The ranges of these parameters (Tab. 2) were chosen to be identical to those that participated in the course of a manual expert calibration for the test watershed. The bounds therefore reproduce the plausible parameter space for this catchment. 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 Freiberger Mulde, a sub-basin of the Elbe River. Before the training, normalization of the data after Eq. (3) 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].

$$x' = \left(x - \bar{x}\right) / \sigma_x \qquad (3)$$

In order to compare the results of the aforementioned Monte-Carlo simulation and the properties of the SOM, seven measures of performance, listed in Tab. 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 2.1. As to the model optimization using the SCE-UA algorithm [Duan et al., 1993] the criterion for successful termination was set to a change of less than 0.05 percent of the performance criterion in three consecutive loops.

2. RESULTS

3.1 Properties of the SOM

After the training each neuron of the 22x15 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. To examine the map the means of different performance measures as well as the mean values of the model parameters on each map element are calculated. 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.

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

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

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

Referring to a) seven performance measures have been calculated for each model realization (Tab. 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. Fig. 2 shows the distribution of the performance measures from Tab. 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 Fig. 2 and Fig. 3 the positions of the neurons remain identical such that each map element refers to identical model realizations in both figures.

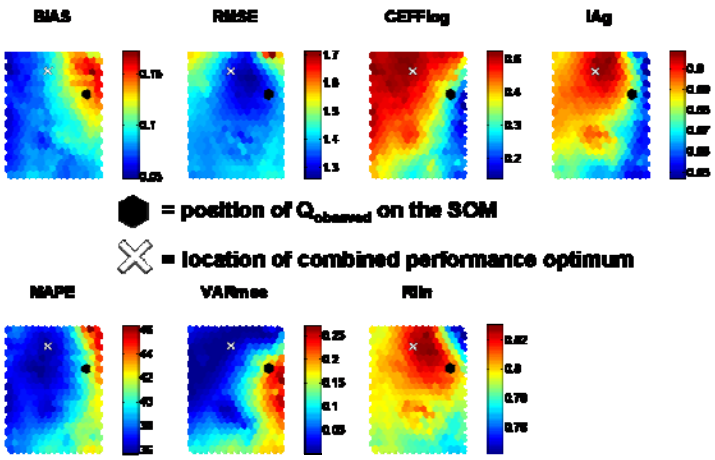

Figure 2. Distribution of the mean values of each performance measure from Table 2 over the SOM lattice, location of the best-matching unit (black dot) for measured discharge time series and common optimum (balance point) of the seven performance measures on the SOM grid (white cross).

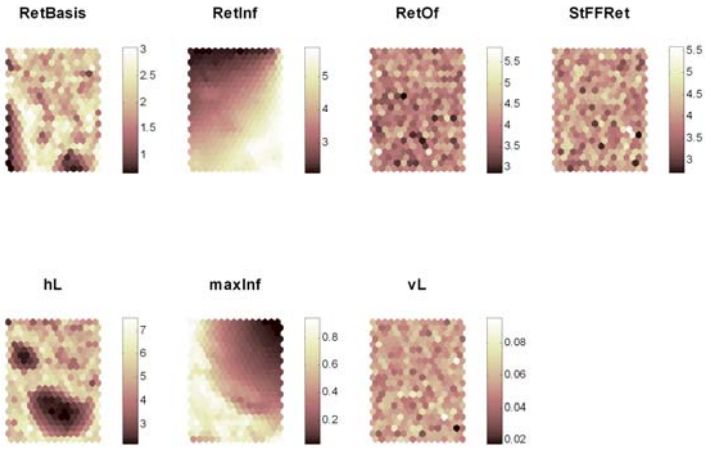

Figure 3. Distribution of the mean values of each model parameter from Table 1 over the SOM lattice.

In Fig. 2 it can be seen that, without providing information on 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. 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. These corroborate the assumption that only the parameters RetInf and maxInf are sensitive. The findings indicate that the parameters RetBasis and hL are subject to interaction with other free parameters, i.e. changes to a parameter influence the operation of another parameter.

3.2 Projection of a reference vector on the SOM

To locate the best-matching unit (BMU) of the measured discharge time series on the map the transformation Eq. (3) is carried out, using the normalization parameters obtained from the input data set, to ascertain that this reference data set can be projected. In Fig. 2 the location of the resulting vector is displayed on top of the performance measure distributions. Additionally, the location of the combined optimum for the seven performance measures is marked, which has been determined as the geometric center of mass of the individual performance measures optima on the map. It can be seen that the position of the BMU neither coincides with any of the expected objective function optima (which are indicated by the colour coding) nor with the common optimum location of the seven performance measures. Tab. 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 the measured time series. By comparing these parameters from the corresponding model runs to the ranges in Tab. 2 it becomes obvious that, with the exception of RetInf and maxInf, all parameter values span the full range of the Monte-Carlo sampling bounds. The resulting model outputs for these 11 realizations is shown in Figure 4c along with the total envelope range of all 4000 simulation outputs in the background and the observed discharge (only the period from 14 January 1995 to 21 October 1995 is reproduced here). It can thus be seen that, compared to the whole set of Monte-Carlo outputs, these realizations obviously comprise a compact subset of "similar" time series.

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

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 (Tab. 1) are shown in Fig. 4b. 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. 4a). Not surprisingly, the SCE algorithm minimizes the RMSE and the same can be stated for most of the remaining performance measures (Tab. 4). The hydrograph corresponding to the optimized model provides a reasonable representation of the measured time series. The recession limbs, however, are more accurately reproduced by the BMU-realizations. This might 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.

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
calibration	0.32	1.58	0.50	0.86	42.36	0.01	0.75
SCE-UA	0.10	1.25	0.49	0.91	36.37	0.06	0.83
SOM	0.13	1.34	0.30	0.88	40.71	0.19	0.81

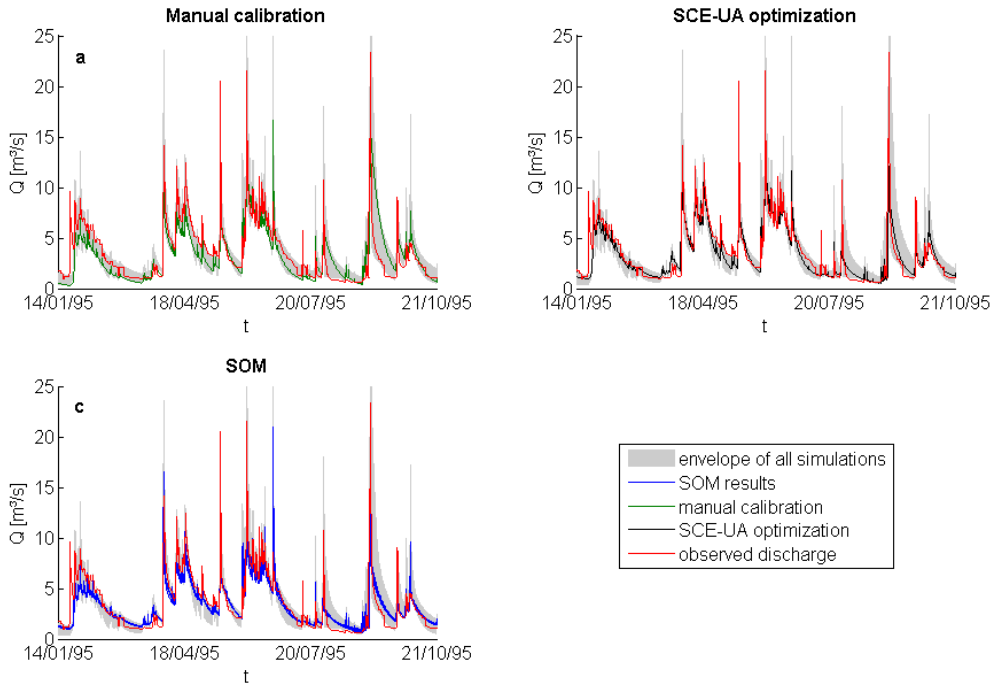

Figure 4. The model realizations as resulting from a) manual calibration, b) optimization with the SCE-UA algorithm and c) the BMU of the SOM for the measured discharge time series. The time-series are compared to the measured discharge and the envelope of the Monte-Carlo simulation.

4. DISCUSSION AND CONCLUSIONS

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. Thus we deduce that the information that can be extracted by these aggregating statistical measures is assimilated and preserved by the SOM. The findings reproduced in Fig. 3, Fig. 4 and Tab. 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. Moreover, 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.

The second experiment 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. Of course, the resolution of the method is dependent upon the number of model time series that participated in the training. The results (Fig. 4c) were achieved with a rather small number of model data items. 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. It is important to notice that the results were achieved without resorting to aggregating statistical measures and therefore, 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.

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 Oliver Buchholz (Hydrotec GmbH) and René Wengel for their invaluable 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)

REFERENCES

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.

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

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.

Herbst, M., and Casper, M. C.: Towards model evaluation and identification using Self-Organizing Maps, Hydrology and Earth System Sciences, 12, 657-667, 2008.

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

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

Lane, S. N.: Assessment of rainfall-runoff models based upon wavelet analysis, Hydrol. 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 Resour. Res., 35, 233-241, 1998WR900018, 1999.

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.: Using SOM in Data Mining, Licentiate's thesis, Helsinki University of Technology, 57 pp., 2000.

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

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.

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

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.