Hydrol. Earth Syst. Sci. Discuss., 4, S2152–S2157, 2008
www.hydrol-earth-syst-sci-discuss.net/4/S2152/2008/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Towards model evaluation and identification using Self-Organizing Maps" *by* M. Herbst and M. C. Casper

M. Herbst

herbstm@uni-trier.de

Received and published: 6 February 2008

The authors greatly acknowledge the very constructive and helpful comments to our paper by the two reviewers which allowed us to prepare an improved final paper.
In the following we comment on the main issues of concern of both reviewers.

Referee 3: Stresses that "the self-organizing map technique, which constitutes the core of the paper, should be explained in greater detail" taking into account that "the typical reader of HESS might not have a strong background on neural networks (...)" and that a conceptual scheme as well as a didactic example "with only 2-4 neurons, and some synthetically generated time series" could help comprehending the technique.

Authors: We fully agree that a good conceptual scheme of the functioning of the SOM technique will considerably improve comprehension and thus the value of the article. We agree as well that it is the author's responsibility to explain all terms that are not commonly known. To this end, an appropriate graphical scheme completes the chapter on the SOM technique in the final paper.
Referring to the text, the authors took great care in providing a complete but very clear description of the algorithm that lists all relevant steps of the algorithm and the properties of the SOM in due detail. The very good and concise description given in Peeters et al. (2007) e.g. shows that a comprehensive outline of the SOM algorithm does not necessarily require much text. A "didactic example" with a SOM consisting of only a few neurons (as can be found in Hsu et al., 2002) does not lend itself to visualize the process of self-organization. The term 'semantic map' describes the self-organizing properties mentioned in the introductory paragraph of page 5 inasmuch as the SOM allocates 'similar' model outputs to the same or nearby locations, i.e. in our study the SOM sorts the model outputs by similarity of the hydrograph pattern. Kohonen (2001, p. 109) dedicates a figure to visualizing speech phonems ordered by similarity. In view of the explanations given in the text we refrained from reproducing this or a similar figure.

Referees: Both referees contend that more details on the Monte-Carlo simulation should be provided. Referee 3 asks whether there is ”a pre-imposed correlation between model parameters” and whether all the parameter combinations are physically plausible.

Authors: The Monte-Carlo simulation conducted for our study uses random sampling of the parameter space. The fixed parameters as well as the ranges of the free parameters (which in the final version will appear in Table 2 instead of Table 1) were deliberately chosen such that they reflect the expert's knowledge on the plausible parameter space for the given test-catchment. On p. 9, lines 23-26 this strategy is explained. Our

HESSD

4, S2152–S2157, 2008

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

aim was to reproduce the situation of a typical model calibration. To this respect, the a priori knowledge on the model parameters has been used. As we basically deal with a conceptual model (where non-uniqueness of the results is an issue) all parameter combinations can be assumed equally plausible. No pre-imposed correlation between model parameters has been assumed.

Referee E.Toth: Has reservations with respect to using the SOM technique for optimization purposes and contends that it ”does not guide the search of the parameter space for identifying the best parameter vectors, like a real optimisation algorithm does.” She further contends that the ”analysis of the simulations issued by the Monte-Carlo approach (...) does not guarantee that the parameter sets that are tested are the best performing” and that even for a very high number of simulations ”the possible parameter sets are obviously much more numerous.” In the same context it is pointed out that the Monte-Carlo simulations shown in Figure 5 were ”probably too good” which is attributed to the choice of the parameter space that was based on the manual expert calibration. It is objected that this would not be ”the way automatic calibrations are implemented.”

Authors: It was not assumed that the parameter sets that were sampled for the Monte-Carlo experiment would be the best performing. It is right that the number (or any number) of sampled parameter sets does not guarantee to include as well the best performing models, which indeed was not the intention of our experiment. The SOM, however, has been used only to select those model realizations from the given set of 4000 Monte-Carlo model results that most closely approximate the pattern of the measured discharge. We agree that this proceeding has very little in common with the functioning of classical optimization algorithms which are based on different species of guided search. Therefore we agree that the term ’optimization’ should not be used in this context unless a procedure is employed that emulates a kind of guided search for the parameter optimum. We propose to use a different terminology in the final version

HESSD

4, S2152–S2157, 2008

of the paper. However, as to the parameter bounds used, also optimization algorithms like the SCE-UA rely on a sensible choice of parameter bounds in order to confine the search to physically realistic or otherwise plausible parameter bounds. In our study the parameter bounds listed in Table 1 have been deemed 'realistic' by expert judgement and thus were deliberately chosen to confine the search. Although we agree that the expert's experience already provided rather good model results. Common parameter bounds were further assumed to be a prerequisite in order to make the SOM and SCE-UA results comparable (though the SOM can only 'choose' from a fixed set of parameter combinations it would have been interesting to find whether the results have something in common).

Referee E. Toth: "Is there a reason for choosing a 22x15 grid?"

Authors: Generally it is left to ones own judgement or experience to choose the dimensioning of the SOM grid. The software used to conduct the experiments offers an option to determine the dimensions of the map using a heuristic algorithm (Vesanto et al., 2000). The number of map units m is first calculated using.

$$m = 5\sqrt{n}$$

with n being the number of data sets used for the training. Further the ratio of the sidelengths is then determined using the two biggest eigenvalues of the covariance matrix of the data. The actual sidelengths are then adjusted such that their product is closest to m. The discretization into 22x15 = 330 groups of time series patterns, as calculated by the algorithm, appeared to be a reasonably fine resolution for our study.

Referee 3: "How is the location of the optimum of the combined performance measure determined" (Figure 4)?

HESSD

4, S2152–S2157, 2008

Interactive Comment

Authors: In the present version of Figure 4 the approximate location of the combined optimum was estimated only. The idea that we followed rather intuitively was to find the location that equals the geometric center of mass of the seven performance optima on the map. All measures are equally weighted with (mass) 1. In the final version of the paper we present an exact approach to calculate the combined optimum of the performance measures on the SOM grid. Figure 4 has been be modified accordingly.

Referee E. Toth: States that more details on the SCE-UA implementation should be added along with ”scatterplots that are more representative of the entire model simulations”, it is further noticed that the results obtained by optimization with the SCE fail to minimize the RMSE (as can be seen in Table 4).

Authors: The unacceptable outcome of the model optimization using SCE-UA has been found to be the result of a flawed configuration of the algorithm.
A new SCE run yielded much more plausible results which are reported in the final version of the paper. The SCE algorithm was run with a maximum of 10000 iterations and 5 complexes with 5 points each. For successful termination a change of less than 0.05 percent of the performance criterion in three consecutive loops was imposed. As expected, the corrected value for the RMSE of the SCE results was smaller than the corresponding values for the BMU results and the manual calibration time series, respectively. The RMSE declined to values equal to the lowest RMSE obtainable with the given set of Monte-Carlo realizations. Also the remaining performance measures as well as the visual aspect of the hydrograph improved accordingly, such that, in terms of objective function values, the result of the SCE optimization outperformed the BMU realizations as well as the manual calibration.
It is important to notice that the BMU realizations, although they were chosen from a comparatively small set of model realizations, already provide a series of hydrographs which, to some extent, still appear to be comparable to the SCE results.
The scatterplots given in Figure 3 (as well as all performance measures) refer to the

HESSD

4, S2152–S2157, 2008

entire simulation run, spanning the period from 1 November 1994 to 28 October 1996. It is not clear to us which type of scatterplot could complement the information of the given figures.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 4, 3953, 2007.

HESSD

4, S2152–S2157, 2008