Self-Organizing Maps as a tool for model time series evaluation

Marcus Herbst[1], Markus Casper[1]

Problem

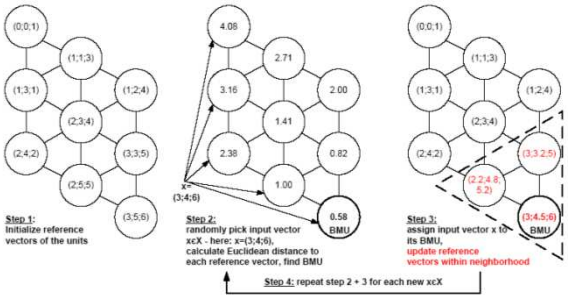
- Aggregating, regression based statistical performance measures have poor discriminatory power because the complexity of the time series features is reduced to a numerical value.
- This implies a shortcoming for applications which depend on the capabilities to differentiate between time series, e.g. model evaluation, optimization and uncertainty analysis.
- Motivation: *Finding a technique to discriminate model realizations without resorting to variance based statistics.*

Approach

- Data: Monte-Carlo simulation with a distributed conceptual rainfall-runoff model on the randomly sampled parameter space is performed (*N*=4000 runs).
- The output time series X are normalized time step-wise to a variance of one and zero mean
- A Self-Organizing Map (SOM), consisting of nodes *i=1…k* which are characterized by a vector

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

with the same dimensionality as the output time series $\boldsymbol{x} \in X$ is trained:



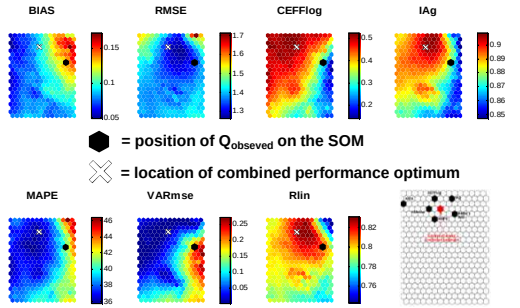
(4) As *k<<N,* each node represents a subset of X

(5) Let *y* be $Q_{observed}$: Find the 'best-matching unit' (BMU) node *c(y)* with reference vector ***mc(y)*** for which

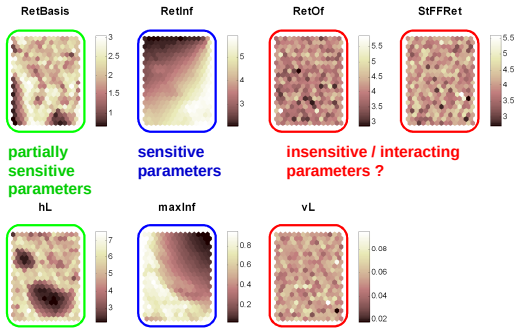
$$\left\| \boldsymbol{y} - \boldsymbol{m}_{c(y)} \right\| = \min_i \left\{ \left\| \boldsymbol{y} - \boldsymbol{m}_i \right\| \right\}$$

Evaluation of the SOM

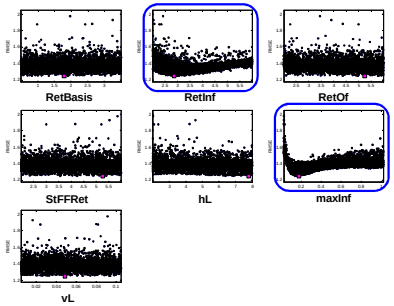
Means of different objective functions calculated for each node, i.e. for each subset of model realizations:



Means of parameter values for each node on the map:

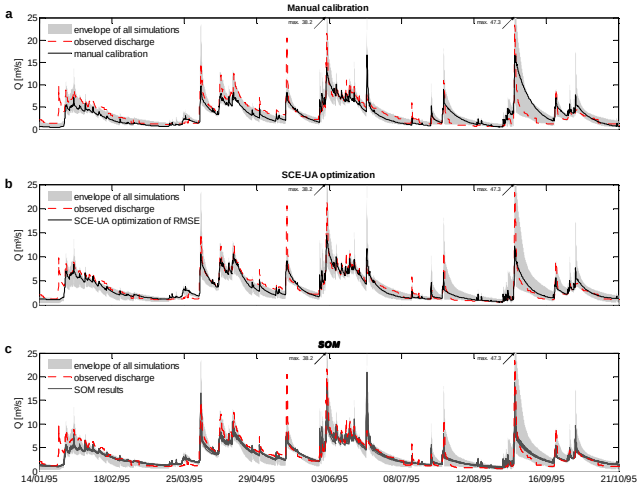


Comparison with parameter scatterplots (RMSE):



Model realizations on BMU for $Q_{observed}$

Comparison of BMU realizations with a) manual expert calibration and b) SCE-UA optimization of RMS-Error:



Conclusions

- The time series on the SOM are topologically ordered by similarity. Simulations can be selected by different properties.
- The SOM is capable of revealing information about parameter behaviour.
- It allows to constrain the ranges of sensitive parameters to a region in the parameter space which represents e.g. $Q_{observed}$.
- Models from the BMU comprise a narrow subset of similar model alternatives that outperform the manual calibration.
- Is capable to account for complexity in time series data.
- Results are likely to improve by a) using larger datasets and b) accounting for redundancies within time-series.

Further Information

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

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)

[1]Physical Geography, University of Trier, 54286 Trier, Germany
e-mail: herbstm@uni-trier.de, Tel.: +49 651 2014555, Fax: +49 651 2013976

Universität Trier