Hydrol. Earth Syst. Sci. Discuss., 5, S2265–S2269, 2009
www.hydrol-earth-syst-sci-discuss.net/5/S2265/2009/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Analysing the temporal dynamics of model performance for hydrological models" *by* D. E. Reusser et al.

E. Pebesma (Referee)

edzer.pebesma@uni-muenster.de

Received and published: 5 January 2009

General comments

The authors address the issue of model performance by analyzing the time structure of model errors, as opposed to lookign at time-aggregated (e.g. averaged) measures of model error. They decided to look at a large number of performance measures, each of them in a time-varying fashion. This creates a huge data set of its own: many performance measures, each varying over time. The authors use self-organizing maps (SOMs), an unsupervised data reduction technique, to analyze structure in this large data set. Synthetic error series with particular characteristics are introduced in the analysis to find out how to interpret the SOMs.

I find the issue of analysing time characteristics of errors is highly relevant. I also find that the approach the authors took–analysing as much as possible measures, then attacking dimensionality problems with unsupervised multivariate techniques–not very illuminating. It creates more obfuscation about what the results exactly mean than enlightment.

Specific comments

1. The authors mention on page 3172 that there are no studies on high resolution temporal dynamics of model performance. One paper that is out there is

   Edzer J Pebesma, Paul Switzer, Keith Loague, 2005. Error analysis for the evaluation of model performance: rainfall-runoff event time series data. Hydrological Processes, Vol. 19, No. 8, pp. 1529-1548.

   this paper looks at structure in error time series, e.g. by analyzing its autocorrelation, and looks at how well errors can be predicted, e.g. by a lagged and/or smoothed versions of model input (rainfall). The paper gives also some relevant references to earlier literature.

2. I find the problem the authors introduce on page 3173, namely that "The large amount of data produced in such an analysis quickly becomes overwhelming and even confusing" a consequence of the author's decision rather than a fact of life. The authors deliberately decide not to choose a single model performance measure, but rather decide to analyse every measure they can find, even time varying, and then they complain that this creates a lot of information that is hard to interpret–hence they need SOMs. I would have found it stronger when the authors had chosen one (or a few) measures that served a particular goal, and had concentrated on that. Which particular goal serves the full collection of performance measures they choose?

3. as the Nash-Sutcliffe measure is widely used and the authors decide to use a

HESSD

5, S2265–S2269, 2009

Interactive Comment

very wide range of measures, it is unclear to me why they used a transformed version of the NS measure, and not both the original one and the transformed one.

4. Table 1 with the list of performance measures raised a few questions; how does NSC measure model error? With t-test, is a paired or a two-sample t-test indicated? MALE–how can you take a log of a negative error? For each of them: are they computed over a time window, and if yes over which window? How does the window size influence the research findings?

5. Table 2: R>0.85 what does that mean? R-squared? or absolute correlation? Or was correlation never strong negative? The 0.85 seems arbitrary (was it important for the analysis?), and weakens the approach. Multivariate statistical techniques are designed to deal with highly correlated data sets, so why the need to preselect?

6. Page 3179: how is the layout of the SOM (x_{max}, y_{max}) chosen, and does it have a consequence for the analysis? Does it matter if the map is square or elongated?

7. The need to apply fuzzy clustering after a SOM makes the whole interpretation very “soft” and hard to follow.

8. The authors acknowledged the R community; a better way of thanking them would be to add the literature references to R and the packages used to the text and references list.

9. page 3194, l 21/21: please point out explicitly a couple of examples of less obvious error types that were overlooked first—the real benefits of this approach.

10. When creating “types of errors”, errors in both ways were created. Eg. Fig 2, subfig 1 shows peak over- and under-shoots. Is it with this method still possible

HESSD

5, S2265–S2269, 2009

Interactive Comment

to classify ("find") a behaviour that is a "peak overshoot", or is a classification only possible as "peak over- or under-shoot"?

Technical corrections the sentence starting at the bottom of 3173 "Classical methods ..." is overly long, unclear and in my opinion not true.

page 3174, line 8: "while preserving the topology" should be "while preserving as much as possible the topology"

page 3174, list item 2: really for each time step or each time window? In case of a window, how large was the window?

page 3174, list item 3: it was not clear to me how exactly highly correlated measures were removed, and which of a pair of correlated measures was removed–details this make the procedure hard to reproduce.

page 3175, list item 4: what is meant by response?

page 3175, "around 20" – why not give the exact number?

page 3176 eq (2)-(4): use indexes in the summation, as in (10) and (11)

Figure 2: subfigure 4 curves seem to be cut off: start x axis at -50?

Figure 5 and 7: use of trellis/lattice graphics could make these collections of graphs more readible, as was done in Fig 2. For figure 5: is it necessary to show the legends for each of the msps?

page 3193, l 21: uncorrelated–are they really uncorrelated?

page 3193 l 24: replace specific with sensitive

page 3194 l 7: "we found 6 classes or clusters"–why 6 and not 4, 5, 7 or 8? Which criterion was used for this?

page 3194, l 21 "The methodology proposed here..."

HESSD

5, S2265–S2269, 2009

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 5, 3169, 2008.

HESSD

5, S2265–S2269, 2009

Interactive Comment