Hydrology and Earth System Sciences Discussions

Interactive comment on "River Flow Forecasting: a Hybrid Model of Self Organizing Maps and Least Square Support Vector Machine" *by* S. Ismail et al.

Anonymous Referee #1

Received and published: 21 December 2010

The basic idea of the paper presented by Ismail et al., using Self Organizing Maps (SOM) as a "pre-processor" for clustering the input data domain in a hybrid modelling approach is not entirely new. Similar ideas have been successfully presented by e.g. Abrahart and See (2000), Moradkhani et al. (2004), Jain and Srinivasulu (2006). However, the combination of SOM with Least Square Support Vector Machines and its application to hydrological forecasting presents a further variant of this successful strategy which is worth being discussed in a scientific paper. Although the paper is factually correct and mostly convincing, with regard to the contents, the paper suffers from very numerous linguistic flaws and some imprecise statements. Maybe there is also room for improvements in some methodical points and the way the comparison of the individual methods is carried out. I would like to recommend this paper for publication after

extensive revision. Detailed comments to the contents are listed below.

1. The authors present a concise review on the applications of SOM in the field of forecasting and modelling. Maybe it would be interesting to mention that SOM can also be used to evaluate model results (see Herbst et al., 2009 and Reusser et al., both in HESS)?

2. p.8188, l.22 contains a very imprecise statement: What could be the reasons for the improvement of LSSVM when the input data domain is subjected to cluster prior to training?

3. I have serious reservations with regard to Eq. 16. This equation requires some explanation, apart from the fact that it constrains the data in a range [0.1+Xmin/1.2Xmax 0.9333], i.e. [0.1 0.9333] assuming that min. discharge would be zero.

4. p. 8191: Three approaches input determination are mentioned in a way that could be more intelligible. In which way are approaches two and three different (l. 8)? Why the lengthy review on input determination methods in Sect. 4.

5. The results of the experiments are compared using only standard statistical measures (MAE, RMSE, R) which is not very informative with regard to the "pros" and "cons" of the individual performances of the three modelling approaches. When does the SOM-LSSVM outperform the other approaches and in which situations doesn't? Maybe the abilities and performances of the three approaches are only problem dependent? Therefore, in addition to Fig. 5, the individual output time series, or parts of it, should be presented in another figure. The authors should in any case provide more insight into the qualities of the results.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 7, 8179, 2010.