Hydrol. Earth Syst. Sci. Discuss., 9, C2273–C2278, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C2273/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Reservoir computing as an alternative to traditional artificial neural networks in rainfall-runoff modelling" *by* N. J. de Vos

Anonymous Referee #1

Received and published: 17 June 2012

This is an interesting and very well-written paper. The topic is also timely; nicely coinciding with the fact that reservoir computing has just been identified as a promising topic for hydrological modelling investigation in a recently published review of neural network river forecasting activities (Abrahart et al., 2012a).

Reservoir computing is a collective term that covers several variants viz. Echo State Network (ESN: Natschläger et al., 2002), Liquid State Machine (LSM: Jaeger, 2001; Lukoševičius and Jaeger, 2009), etc. However, given that only ESN modelling is compared and contrasted against other related methodologies in the reported analysis,

perhaps the final paper might be better entitled "Echo state networks as an alternative to traditional artificial neural networks in rainfall-runoff modelling".

I am somewhat uncertain as to why the present author has opted to restrict his river forecasting analysis to a simple consideration of daily one-step-ahead predictions, since this situation represents the least challenging of all potential hydrological modelling opportunities that could have been pursued. More context and justification is necessary to explain why such a problem was selected for investigation in the first place and the potential relevance of any identified findings. The reported analysis whilst interesting may indeed serve no strong scientific or practical or operational purpose? Surely the main point of a demonstration project, such as the one which is being reported, should be to showcase the numerous strengths and weaknesses of a particular algorithm, by providing a rigorous assessment, performed against a set of increasingly more demanding requirements and/or complex numerical explorations?

The only other known hydrological modelling paper involving reservoir computing is that of Coulibaly (2010) on forecasting monthly water levels for the Great Lakes. He also used an ESN. That paper was subsequently discussed and extended by means of a simple linear benchmarking operation in Abrahart et al. (2012b). The current author has not identified or included a consideration of the latter publication in his opening paragraphs and is accordingly directed to it for additional argument. The principal concern in that initial study was a need for more accurate longer term forecasts i.e. greater than one-step-ahead. ESN modelling was found to be substantially superior over longer lead times and this appeared to be its greatest potential offering. The current paper is clearly not fully testing or highlighting what might indeed prove to be its best advantages: although pointers to further research are provided in the closing paragraph. Stronger engagement with published material is called for.

The discussion section on recurrent and partial recurrent neural networks could be improved by a more detailed clarification of terminology regarding the different architectural arrangements, perhaps supported by a hierarchical schematic. The basic structure of a fully recurrent model is a network of neuron units, each with a directed connection to every other unit. Any other neural network variant should be classified as either a partial recurrent network, or a feedforward network, according to the permitted direction(s) of information flow and/or which particular components are allowed to be connected.

In data-driven modelling the data is all-important. I would have expected to see a set of tabulated statistical descriptions covering all datasets and subsets that are used in a reported investigation. Simply referring the reader to an earlier paper, published by a different author in a different journal, is not good practice since each individual paper should contain sufficient information within its pages to support the production of a full peer-reviewed publication as a stand-alone entity in its own right.

In many of the reported instances it is apparent that persistence and/or linear benchmarking models do reasonably well in comparison to some of their more complicated neural network counterparts, suggesting that the matter under examination is in several cases perhaps being seen as either a near-linear or perhaps marginally non-linear problem (Abrahart and See, 2007). Full particulars on the linear correlation analysis and average mutual information testing, conducted between each input and output series, must as a result be provided since such mechanistic selection procedures could be a significant controlling factor. This is particularly important in cases where a high degree of near-linear modelling is apparent since the input selection process could perhaps be introducing a bias effect. The data was first converted into a normalised format but thereafter apparently pre-processed using principal component analysis. I do not understand exactly what has happened in the latter process or why it was necessary. Further clarification is required.

I am slightly confused about the reported use of training and cross-validation datasets in the modelling process. Most models were calibrated on the training dataset, with the cross-validation dataset being used to perform early stopping. This is standard practice in the field. It means that two independent datasets were included in the model devel-

opment process and the second dataset may in fact have actually handicapped the production of superior solutions, as opposed to providing a set of clear improvements related to enhanced model generalisation/ prevention of overfitting. Early stopping was not used in the linear regression modelling or reservoir computing operations so were these particular models calibrated on just the training dataset, or a combination of training and cross-validation datasets? If different development datasets are used, how is inter-modelling fairness achieved, given that particular datasets will offer different modelling advantages and shortfalls. How is everything balanced out in such cases?

In situations where the last known discharge record is included as a predictor in the modelling process, a neural network model will tend to become a “prisoner of that measurement”. This issue is perhaps best exemplified in recent attempts to identify and specifically fix such problems by Abrahart et al. (2007). Their paper should be cited and included in the list of referenced material.

I wonder if it would have been more logical to include a hyperbolic tangent transfer function in the output neurons of the standard neural network models, so as to match the fully recurrent method?

The model comparison section is rather limited. The author states that the overall performance of all models is quite good in comparison to various conceptual models presented in a different paper applied to the same dataset(s). Surely some sort of analytical comparison should be provided? I am also concerned about the fact that the final modelling comparison is primarily restricted to a consideration of statistical metrics and no hydrographs or scatter plots are depicted or inspected. This would of course enable a more detailed analysis of modelling outputs to be performed from which a deeper understanding of matters might be obtained.

The author has included two additional variants of reservoir computing in the final stages of his paper which appear to be an afterthought. It would be better if these items were considered as individual stand-alone models and included in the opening sections as alternative solutions, under the guise of some larger overall predetermined analytical operation. If not, it raises the question, as to whether or not some of the other models under test could also have been improved following a detailed inspection of their respective difficulties and failings?

Table 2: please explain the difference between trained and untrained weights.

Figure 1: more detailed explanation required for error loop components.

Figures 4, 5 and 6: the plots are deceptive since each graphic is drawn to a different scale and so one cannot compare the different basins in a meaningful manner. Please ensure that all plots are drawn to the same vertical scale to support improved reader interpretation and prevent misunderstandings. There is no legend. I can only assume that multiple runs were performed on each different type of model and that the red and blue represent some sort of mean and standard deviation values? If so, how many model runs were performed? The main text must be amended to include an explanation for this missing aspect of your overall modelling methodology.

References

Abrahart, R.J., Anctil, F., Coulibaly, P., Dawson, C.W., Mount, N.J., See, L.M., Shamseldin, A.Y., Solomatine, D.P., Toth, E. and Wilby, R.L. (2012a) Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Progress in Physical Geography [posted online, ahead of print: 14 Jun '12]. doi:10.1177/0309133312444943

Abrahart, R.J., Heppenstall, A.J. and See, L.M. (2007) Timing error correction procedure applied to neural network rainfall-runoff modelling. Hydrological Sciences Journal 52: 414-431. doi: 10.1623/hysj.52.3.414

Abrahart, R.J., Mount, N.J. and Shamseldin, A.Y. (2012b) Discussion of “Reservoir Computing approach to Great Lakes water level forecasting” by P. Coulibaly [J. Hydrol. 381 (2010) 76-88]. Journal of Hydrology 422-423: 76-80. doi:

10.1016/j.jhydrol.2011.10.006

Abrahart, R.J. and See, L.M. (2007) Neural network modelling of non-linear hydrological relationships. Hydrology and Earth System Sciences 11(5): 1563-1579. doi: 10.5194/hess-11-1563-2007

Coulibaly, P. (2010) Reservoir Computing approach to Great Lakes water level forecasting. Journal of Hydrology 381(1-2): 76 – 88.

Jaeger, H. (2001) The "echo state” approach to analysing and training recurrent neural networks. German National Research Center for Information Technology, GMD Report 148.

Lukoševičius, M. and Jaeger, H. (2009) Reservoir computing approaches to recurrent neural network training. Computer Science Review 3(3): 127–149. doi: 10.1016/j.cosrev.2009.03.005

Natschläger, T., Maass, W. and Markram, H. (2002) The "Liquid Computer": A novel strategy for real-time computing on time series. Telematik 8(1): 39–43.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 6101, 2012.