Hydrol. Earth Syst. Sci. Discuss., 11, C2668–C2671, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C2668/2014/
© Author(s) 2014. This work is distributed under
the Creative Commons Attribute 3.0 License.

Hydrology and Earth System Sciences
Discussions

Interactive comment on "A strategy to overcome adverse effects of autoregressive updating of streamflow predictions" *by* M. Li et al.

M. Thyer (Referee)

mark.thyer@adelaide.edu.au

Received and published: 24 July 2014

General Comments

This paper presents a modification to existing approaches for handling autoregressive errors in streamflow modelling in a forecasting context. I applaud this paper for undergoing a detailed analysis of the issues that are encountered when endeavouring to deal with both heteroscedasticity and autocorrelation in hydrological modelling errors. Something which we think should be straightforward, but is actually quite challenging to get right. The paper is fairly well written, but needs some improvement (see minor issues). The results presented, while quite promising, are currently not sufficiently convincing to warrant publication. Please see the list of major issues below. These issues

need to be addressed prior to publication.

Major Issues

1. More metrics are required to verify performance

Currently the three methods, AR-Norm, AR-Raw and RAR-Norm are evaluated by visual inspection of a few events and using the NSE as an evaluation criteria. A wider range of metrics is needed. In a forecasting context, it is not simply the NSE which is used to evaluate predictions, users are also interested in the statistical properties of the predictive streamflow distribution, such as reliability and precision. It is common for these metrics to also trade-off against one another, so it would be interesting to see if that occurs in this case. Furthermore, the NSE is heavily weighted towards better predictions of high flows. It is recommended that authors use metrics that evaluate the full predictive streamflow distribution and use precision and reliability metrics, such as they have used in past, e.g. Wang and Robertson [2011] or see for example Evin et al. [2014].

2. Robustness of the results with respect to the hydrological model

Line 20 page 6044 makes the point that AR-Raw performs better than AR-Norm and state "this suggest that more robust performance can be expected of base hydrological models with AR models are applied to raw errors". Sectio, 4.2 is devoted to discusses that the AR-Norm model, produce poor performance of the hydrological model. However, this is based only a single hydrological model, GR4J. When Evin et al. [2014] applied an equivalent to the AR-Norm model (but with linear heteroscedatic errors, rather than log-sinh transformed) to the 12 MOPEX catchments they found similar poor model performance for GR4J for some catchments, but this did not occur when the HBV model was applied. This provides strong evidence that the problems with AR-Norm is not necessarily generic, but hydrological model specific. It is recommended that the authors trial a different hydrological model, e.g. HBV, and see if the results are similar. If they are, then this provides a greater robustness of the model results, and

HESSD

11, C2668–C2671, 2014

greater confidence for the hydrological community to adopt this method.

Providing more metrics with a wider range of hydrological models would be better test the extent of the problems with AR-Raw and AR-Norm and the robustness of the results. For example, Figure 3, shows the error over-correction problem with AR-Norm occurs in only 10-20% of cases, which is not very high. Given also that the poor performance of the AR-Norm method is hydrological model specific, further testing and metrics are required to verify the robustness of the proposed approach.

3. Ability to compare results with previous studies.

This is more a general comment of an issue which is a common blight for the progress of the hydrological scientific community. One of the big challenges for reviewers (and readers in general) is the ability to compare results between different studies, due to differences in implementation. As an example, Evin et al. [2013] showed that the equivalent to AR-norm was better than AR-Raw, while Evin et al. [2014] showed that AR-Norm can degrade hydrological model performance for GRJ, but not HBV. While Schaefli et al. [2007] showed that AR on raw errors lead to better inference, while this study showed a AR-hybrid (norm and raw) (see minor comment 3) works better than both AR-Norm and AR-Raw. However in all these studies, there are differences in their approach and case study application. For example, Evin et al. [2013,2014] used a linear heteroscedastic residual error model, Schaefli et al. [2007] used a mixture of Gaussians for their error model, while this study used a log-sinh transformation with modification for zero flow occurrences. Furthermore, each study had a different set of case study catchments. It concerns me that the conclusions of each of these studies could be sensitive to these differences rather than differences in the way the AR is handled, and it makes it very difficult for hydrological science to move forward. This is the reason why Evin et al. [2014] choose to use the MOPEX dataset, as it least provides a common set of catchments to previous studies. I would suggest to these authors to include the 12 MOPEX catchments as used by Evin et al. [2014] to enable better comparison. This is not an essential criteria, but it would increase the ability to

HESSD

11, C2668–C2671, 2014

compare the results, and test its compare robustness against previous results.

Minor Issues

5. Page 6039 Line 20-25. The assertion that these equations represent the median needs further derivation (perhaps in an appendix), as it is not clear to me. For example, the error term e(t) is completely dropped from eqs 4 and 5. This assumes that median of Z-1(et)=0, now median(et)=0, but, I'm not convinced that median of Z-1(et)=0, due to the use of the log sinh transformation which takes into account zero flow occurrences.

6. Page 6045, Eq(8). It is very confusing using the subscript (R) for both AR-Raw and AR-Norm. Please use a different subscript for RAR-Model

7. RAR model is essentially a hybrid of AR-Norm when it over-corrections, use AR-Raw. Suggest to change name of RAR_Norm to AR-hybrid. Also, why did the authors choose not use the phi term, i.e. Q(s,t)+ phi*[Q(t-1) – Q(s,t-1)] in last line of eq 8. Some justification of this is needed.

8. Figure 3 – Q(M,t) is used before it is defined. Please define it earlier in the manuscript.

9. Agree with B. Schaefli, the superscript notation is hard to read. Please change to increase readability

10. Agree with B. Schaefli, re structure, the new method RAR should be presented in Section 2. All methods should be in a method section, all results in a results section

11. Please also provide details on the algorithm used to maximize the likelihood – was it SCE or something else?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 6035, 2014.