Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2017-9-RC3, 2017

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Bayesian joint inference of hydrological and generalized error models with the enforcement of Total Laws" *by* Mario R. Hernández-López and Félix Francés

Anonymous Referee #1

Received and published: 26 February 2017

I am afraid I have to firmly disagree with the majority of these arguments. I will more succinctly summarize the key concerns raised in my original review, most notably the un-substantiated conclusions.

The current manuscript clearly uses the findings of Schoups and Vrugt (2010) and Evin et al (2013, 2014) as a motivation - this is clear from the first line of conclusions and from the introduction, and generally throughout the manuscript. So what are the problems raised by SV10 and E13/E14 that have been resolved in the current manuscript?

The publications by SV10 and E13/E14 empirically found that joint inference (WLS-AR1 scheme) that includes the autocorrelation parameter 'rho' is not always stable - it performs well in catchments such as the French Broad River, but falls apart in many

other catchments such as San Marcos, etc. And the manifestation of the problem is explosive behavior of the prediction limits - an issue of clear concern to a modeller. As an alternative to joint inference, E14 offer a post-processing method, WLS-AR1-PP, to avoid the instability problem and produce well-behaved prediction limits, including in the problematic catchments such as San Marcos.

If the manuscript under review is to claim to resolve the problems with joint inference raised by SV10 and E13/E14, it should clearly show a joint-inference technique that produces well-behaved prediction limits in the problematique catchments, with performance at least as good or better than, eg, the post-processing method WLS-AR1-PP from E14. Otherwise where is the actual improvement???

This request is especially salient in view of the authors response to Comment #2 (top of page 2 of response), where they state that "The cause of this 'non-robustness' (reported by Evin et al 2014) is not the possible interaction among parameters. The cause is a statistical one: the non-fulfillment of Total Laws. All other detected problematic (or not) effects, as the possible interactions, are a consequence of this severe statistical mistake."

There is clearly no evidence shown by the current manuscript in support of this suggestion: Evin et al 2014 reported good performance for the WLS-AR1 scheme in the French catchment (and a few others), and reported non-robustness in catchments such as San Marcos (and a few others). The present manuscript only considered the French Broad River.

Therefore one of the manuscript's central assertions fails on at least three counts:

1) There is no evidence given in the manuscript that fulfilling the TVL avoids non-robustness (this requires an application in the problematic catchments such as San Marcos);

2) If violation of TVL is such a severe mistake, why does it not manifest itself in the

application of the WLS-AR1 scheme in half of the Mopex catchments (eg, in the French Broad), and

3) The WLS-AR1-PP scheme does not satisfy the TVL conditions either, yet does not exhibit non-robustness in *any* of the 12 catchments tested by E14.

The manuscript methodology is just not setup to answer any of these questions, which precludes any empirical support for most of its arguments - most notably, with respect to the practical importance of the TVL conditions.

If I read the authors response to the review correctly (comment #6), they consider the fulfillment of TVL more important than properties such as reliability and precision of the predictive distribution. I am a little puzzled by this perspective - if a statistical scheme that satisfies the TVL conditions produces predictions that are no better than predictions obtained when violating the TVL, according to metrics such reliability and precision (which are the key metrics used in the forecasting community), then it suggests that fullfilling TVL does not bring much practical value. Simply fulfilling the TVL conditions can not be considered a *practical* advance in its own right until it is shown to be beneficial in some *practical* performance aspect. This is especially so if enforcing the TVL conditions sacrifices other performance criteria of already-established value to hydrological forecasters and other modellers.

The other points raised in the response - such as that WLS does not capture the error time dependence - have already been resolved in previous work such as SV10 and E13/E14. The methods recommended in these published studies already incorporate AR1 assumptions and in terms of error time dependence the current manuscript does not offer anything beyond the AR1 treatment.

In summary, I am not at all convinced by the authors response - if they wish to claim to have solved the *problem* with joint inference including 'rho', they should demonstrate their technique on a catchment where existing joint inference techniques fail. The current manuscript simply shows a new, more complex technique that, according

to common performance metrics, performs no better than existing techniques on the French Broad river, which is one of the most well-behaved catchments to appear in the hydrological literature, where the majority of hydrological models and error models already perform very well. I do not believe this constitutes a substantiated contribution at the level of a peer-reviewed publication in the hydrological community.

This is not to discount the *potential* contribution of the authors' methods, just that, as it stands, the case study setup precludes any meaningful claims of practical improvement over existing techniques.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-9, 2017.