Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-635-RC1, 2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across a large-sample of catchments in Great Britain" *by* Rosanna A. Lane et al.

Thibault Mathevet (Referee)

thibault.mathevet@gmail.com

Received and published: 17 February 2019

I carefully read the paper by Lane et al.. This paper appeared to be particularly clear, well written and easy to follow. Scope and objectives are stated clearly, the presentation of results is rather straightforward. As you probably know it, I appreciate this kind of study on a large sample of watersheds. I am very happy to know that such a large sample exists for GB. Studies on large sample give generality and robustness to the results. This paper gives insights on the general hydrology of GB and predictive capabilities of 4 simple rainfall-runoff models. I really appreciated §4 and §5, par-

ticularly analyses linked to the seasonality (fig 4), BFI (fig 5, 6), and water balance closure (fig 7). Thanks to this large sample of watersheds in GB with a variety of hydrologic/hydrogeologic functioning (even in the same country), these results appear to be robusts, with a general interest. The link between BFI (main undergroud processes) and model structure agility is really interesting.

»> Main comments Evaluation of model performance and selection of model :

Authors decided to use the classical Nash-Sutcliffe efficiency (NSE) index to evaluate model performances (and select behavioural models, NSE > 0.5). NSE index is famous and widely used in Rainfall-Runoff modeling. Even if the perfect efficiency index do not exists, this index is also known to have some drawbacks (Schaefli and Gupta, 2007, among many references). Gupta et al. (2009) introduced the Kling-Gupta efficiency index that allows to explicitly account for bias (mean and variability) and correlation, in the evaluation of model performances. Given the ambition of this paper, I would recommand the authors to consider in their analyses the Kling-Gupta efficiency index, or at least to decompose their results in terms of correlation and mean bias.

Poor performances on floods :

Authors found that the different models had poor performances on floods, which is generally the case when classical modeling schemes are used to optimise or select parameter sets. I appreciate the simple way authors evaluate models on flood values, however I would add a figure to explain the two metrics. One of the main drawback of the NSE (and linear regression as well) is that the standard deviation of the simulated time-series is biased and underestimated, i.e. flood underestimated and drought overestimated. Among other arguments, this drawback partly explain why flood values are underestimated. I would add at least a comment on the fact that this statement is dependent on the behavioural model selection metrics in §5.4. If authors update their paper using KGE to select their behavioural models, they might revise (a bit) their findings on model performances for floods.

Focus on droughts ? :

Given the ambition of this paper, I think that this paper would also benefit from a focus on droughts. Hence, analyses on doughts could be complementary to analyses on relative model performances (among the 4 tested structures), since droughts might also be driven by BFI in GB ? The link with groundwater flows could also be shown, if a focus on droughts is done. Authors could use the same metrics as for floods. It could be better to use the 10 days or 30 days annual minimal value, instead of the annual minimal value, which could be highly impacted and uncertain.

»> Minor comments :

In §1.1 : authors discuss the benefits of national scale hydrological modelling. Another benefits could be the production of parameter libraries, which could be used for regional studies or model calibration on poorly gauged to ungauged basins or engineering studies. Authors can make references to papers on this subject (Perrin et al., 2008 ; Rojas-Serna et al., 2016 ; or some other works by Seibert).

In §5.1 : authors did not use a snow accumulation and melt routine in their modeling framework. Very simple snow routine are availables, in the spirit of the simple models proposed in FUSE. The CemaNeige routine could be a good candidate to improve model simulations on the few catchments where it's necessary. Depending on the proportion of snow impacted catchments, using a snow routine would improve model performances and the paper, as it could give answers to some hypotheses of the paper. Valéry, A., Andréassian, V., Perrin, C., 2014. 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 1 – Comparison of six snow accounting routines on 380 catchments, Journal of Hydrology, 517(0): 1166-1175.

Valéry, A., Andréassian, V., Perrin, C., 2014. 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments, Journal of

Hydrology, 517(0): 1176-1187.

In §5.3 : authors discuss about groudwater flows between catchments, with losses or gain of waters. This problem is not new and some conceptual modelisation could be found in the litterature since one or two decades. In a natural context, authors could make a reference to Le Moine at al. (2007, 2008) papers about groundwater flows and water balance closure. The existence of such groundwater flows in permeable geological context (chalk, limestones and/or karstic systems, etc.) was one of the reasons of the development of a groundwater exchange function within the GR model family. The use of this function should be motivated by (hydrogeologic) evidences of such groundwater flows (in order to avoid "overfitting" of the water balance, i.e. fudge factor), but might be usefull in catchments where water balance is difficult to close, such as the one influenced by chalk aquifers in southeast england.

Le Moine, N., Andréassian, V., Perrin, C. & Michel, C., 2007. How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. Water Resources Research 43(6), W06428

Le Moine, N., V. Andréassian, T. Mathevet, 2008. Confronting surface- and groundwater balances on the La Rochefoucauld-Touvre karstic system (Charente, France), Water Resour. Res., 44, W03403

»> Last comments :

P4, l22 : I would also make a reference to Perrin et al. 2001 here

P7, l20 : mistake with O (mean of observed discharge)

P10, l22 : values instead of vales

P17, l8 : for catchments, repeated 2 times

In §2, I would give an estimation of the proportion of watersheds where snowmelt processes are observable (solid precipitation >20% of total precipitation ?)

Table 1 is not cited within §2

In §3.3, the +/- 13% concerning streamflow uncertainties for flood should be a bit more explained. To which probability range this uncertainty refers ? Is it one or two standard deviation (or something else) ?

In Figure 2, I would put the number of free parameters to calibrate.

Conclusion : I really appreciated this paper and I would like to congratulate the authors for their work.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-635, 2019.