Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-635-AC1, 2019 © Author(s) 2019. This work is distributed under the Creative Commons Attribution 4.0 License.



HESSD

Interactive comment

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.

### Rosanna A. Lane et al.

r.a.lane@bristol.ac.uk

Received and published: 26 April 2019

We thank Thibault Mathevet for taking the time to review our manuscript, and for his helpful comments. Our responses to each individual comment are outlined in bold below.

Printer-friendly version



### Summary

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, particularly 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 underground processes) and model structure agility is really interesting.

Response: We thank Thibault Mathevet for taking the time to thoroughly review our manuscript, and for his positive comments.

#### 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

# HESSD

Interactive comment

**Printer-friendly version** 



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.

Response: The NSE index was chosen for this analysis as it is so widely used and easy to interpret. Given our focus on floods, it is also a good choice as it emphasizes the fit to peaks more than KGE. However, we agree that there are drawbacks to only using the NSE index and so following this comment, we will provide additional analysis looking at the correlation, variance 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.

Response: Thank you for pointing this out. In response to this comment, we will ensure the flood metrics are explained as simply as possible and consider adding a figure to help explain how they were calculated. We selected NSE as it emphasizes the fit to peaks, whilst KGE is more general, but we acknowledge that it has drawbacks and no global performance measure is useful in all situations, especially when looking at extremes. We will therefore keep using NSE to select behavioural models, but as suggested we will comment in the discussion on how behavioural model selection metrics influences estimation of flood values.

## HESSD

Interactive comment

Printer-friendly version



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 droughts 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.

Response: We agree that focusing on droughts could be an interesting question in itself, however we feel that it is out of scope for this paper. We are aiming to give a general overview of the capability of models, with a focus on high flows. Drought and very low flows is a more complex problem to address, and more likely to be influenced by human impacts in managed catchments. We therefore think adding this would be too much for one paper.

#### 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).

Response: Thank you for this idea, we will include discussion of this benefit in the manuscript, and ensure best-performing parameter sets are made available through a DOI or supplementary information.

## HESSD

Interactive comment

**Printer-friendly version** 



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.

Response: Thank you for this comment, but we do not think it would be feasible to run all simulations again with a snow routine. We originally decided not to use a snow routine as only a relatively few catchments were snow impacted. To check this, we have calculated snow fractions for all catchments, as the sum of the rainfall on days when daily mean temperature is less than 0 degrees Celsius divided by the total sum of the rainfall for the whole time period. This confirms that only a small proportion of catchments are snow impacted (13 catchments out of the 1127 have a snow fraction of more than 10%, and no catchments have a snow fraction of more than 17%). As the concept of the paper is focused on benchmarking the capability of these lumped models, and not model development, we feel that addition of a snow routine is out of scope.

In §5.3 : authors discuss about groundwater flows between catchments, with losses or gain of waters. This problem is not new and some conceptual modelisation could be found in the literature 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

# HESSD

Interactive comment

Printer-friendly version



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 useful in catchments where water balance is difficult to close, such as the one influenced by chalk aquifers in southeast england.

Response: Thank you for highlighting these interesting and very relevant papers. We will add this into the discussion section where relevant. However we have not yet done a comprehensive analyses of gaining and losing streams in the UK aquifer systems. This is indeed research that our group is currently conducting in more detail (separate PhD on improving ground water representation in models). Certainly from our preliminary analyses it is very difficult to attribute these losses and gains.

Last comments

P4, I22 : I would also make a reference to Perrin et al. 2001 here **Response: We agree this is a relevant paper, it has been added.** 

P7, I20 : mistake with O (mean of observed discharge) Response: Thank you for noticing this, it has been corrected.

P10, l22 : values instead of vales **Response: This has been corrected.** 

P17, l8 : for catchments, repeated 2 times **Response: The repetition has been removed.** 

HESSD

Interactive comment

**Printer-friendly version** 



In §2, I would give an estimation of the proportion of watersheds where snowmelt processes are observable (solid precipitation is more than 20% of total precipitation ?) **Response: We agree that this would be useful and will add a map of snow fractions to figure 1 (as discussed above).** 

Table 1 is not cited within §2

Response: Thank you for spotting this, we have now added the citation: "The catchments cover all regions and include a wide variety of catchment characteristics including topography, geology and climate (see Table 1)."

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) ?

Response: The +/-13% represents the 95th percentile range of the discharge uncertainty bounds and was chosen as a representative discharge uncertainty for annual maximum flows from a national analysis of discharge uncertainties (Coxon et al, 2015). We will better clarify this in the text.

In Figure 2, I would put the number of free parameters to calibrate. **Response: These will be added to the figure caption.** 

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

Response: We would like to thank Thibault Mathevet for his thorough review, and useful suggestions for the manuscript.

**HESSD** 

Interactive comment

Printer-friendly version



