Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-3-AC2, 2019

Hydrology and Earth System Sciences Discussions

HESSD

Interactive comment

Interactive comment on "A Multi-Objective Ensemble Approach to Hydrological Modelling in the UK: An Application to Historic Drought Reconstruction" *by* K. A. Smith et al.

K. A. Smith et al.

k.a.smith@ceh.ac.uk

Received and published: 29 April 2019

Authors Response to Anonymous Referee 2

Authors comments are given in a bold typeface

This paper uses a multi-objective approach to calibrate a fairly simple hydrologic model to predict discharge at a large number of catchments in the UK based on precipitation and temperature observations. The stated purpose of the exercise was to hindcast streamflow during historical early 20th century droughts that occurred prior to the systematic collection of discharge observations on UK streams, but (crucially) not before available meteorological records. The results show that the relatively simple hydrologic model that was used (4 parameters) was able to capture streamflow variability well, over the wide range of catchments included in the survey. The study showed little evidence of non-stationarity in parameter calibration, which allowed historical droughts to be hindcasted with a decent level of confidence.

Major remarks

The study is methodologically solid. The paper is well written and methods and results are described clearly and in sufficient details. However, I am not sure I understand the contribution of the paper beyond a solid regional study of UK streams. This is without a doubt a useful practical contribution for the UK water resources community, but you should do a better job at discussing general implications of the research in the introduction and discussion. To be excessively blunt, as a scientist that has no particular interest in UK streams (like a large chunk of HESS readership), why should I care? To be a bit more specific, you explicitly lists the intended contributions of the paper in the conclusions (L527). At face value, these contributions are sufficiently general to interest non-UK readers and should be stated upfront (the intro is very much UK specific currently). However, I think that these arguments currently lack substance and should be further developed:

RESPONSE: We thank you for your comments, and appreciate that the introduction could be better framed. We believe that the methods employed in this study are applicable elsewhere across the globe, as well as in time. The multi-objective approach to model calibration used here is not exclusive to the UK, nor to reconstructions, but may also be used to calibrate models elsewhere for flow forecasting and longer term projections. Similarly, it could, with sufficient computational resources, be applied to more complex hydrological models. Furthermore, we believe that the data produced from this research will be of wider interest in the framing of historic flows and extreme events from a European perspective. If you agree that the contributions outlined in the discussion are of

sufficient interest to wider readers, we will revise the manuscript to make these points clearer in the abstract and the introduction.

1. You mention your multi-objective calibration approach as the first general contribution of the paper. As you admit yourself (L91), the concept itself of multi-objective calibration is not new and the section where you describe model selection (3.4) is particularly cryptic. If multi-objective calibration is indeed a key contribution of the paper, please describe the approach specifically (How are the model parametrizations "ranked"? How are each of the criteria weighted to come up with a composite ranking?) and spell out clearly what the novelty is compared to existing approaches.

RESPONSE: We apologise that the method has not been clearly set out, and that you found section 3.4 cryptic; we will endeavour to make it more transparent. We will likely include the code that was used for the ranking process in the supplementary materials for the readers' reference. The third reviewer has also commented that we need to put our method in the context of existing multi-objective calibration approaches, so we will make sure this issue is addressed in the revised manuscript.

2. Second, you claim that the approach can be used not only to hindcast droughts but also to predict catchment responses to future climate change. In order to make such a claim, you ought to address the elephant in the room, which is that your approach does not accommodate non-stationarities in the calibrated parameters (e.g., related to land use change and human adaptation). Your result suggest that these factor were not much of a problem for historical simulations (except for heavily altered catchment), but if there is one thing that climate studies tell us is that the past is not necessarily representative of the future. I do agree that your results are interesting and can be leveraged to study the hydrological impacts of climate change, but the implied caveats and potential avenues to go around them should be discussed. I am specifically thinking of the potential to leverage satellite observations of land use change and/or modules integrating human adaptation to large scale hydrological models (e.g, Bierkens 2015, Calvin

Printer-friendly version

Discussion paper

2018).

RESPONSE: We agree that land use changes and human adaptations are likely to influence flows significantly in the context of climate change projections. However, we are reassured by the integrity of the model results when compared to the longer observed time series. Previous modelling studies have used lumped catchment models to simulate flows under climate change (e.g. future flows hydrology, Haxton et al 2012), and the results have been widely employed in water resource management simulations. We anticipate that this modelling framework, applied to more recent climate projections such as UKCP18 may be equally useful for decision makers, especially in the near-natural low flows benchmark network catchments, where water resource managers may use the flow projections to assess water availability, and subsequently run the flow projections through water resource models to simulate the impacts of changes in human influence over time. We discuss the lack of human influence in the model in the discussion section, but we will add this caveat to the mention of future applications, and also reference the Future Flows Hydrology study in the manuscript.

Haxton, T.; Crooks, S.; Jackson, C.R.; Barkwith, A.K.A.P.; Kelvin, J.; Williamson, J.; Mackay, J.D.; Wang, L.; Davies, H.; Young, A.; Prudhomme, C. (2012). Future flows hydrology data. NERC Environmental Information Data Centre. https://doi.org/10.5285/f3723162-4fed-4d9d-92c6-dd17412fa37b

3. Third, you argue that the study provides important spatio-temporal data on historical drought in the UK (so far so good) which can be used to plan and forecast the onset, duration and termination of drought events in the UK and overseas. First off, it is not clear to me how, specifically, how the historical reanalysis you describe can be used to forecast and mitigate the effect of future droughts (see previous point) – if you have a specific idea here, please make it explicit.

RESPONSE: Historical data can provide vital context when faced with an on-

going drought episode. Whilst, as you say, the past may not necessarily be representative of the future, using ensembles of historical drought events can gain insight into the probabilities of the termination of a current event over a certain time period (e.g. Parry et al, 2018). Knowledge of historic events can also be used to explore statistical correlations with atmospheric drivers of droughts that may help predict the onset of events (e.g. Lavers et al, 2015). In these approaches, extending the hydrological record by 70 years significantly increases the sample of historic drought events from which to conduct such research. Furthermore, the modelled data may be used to extend streamflow records used in seasonal hydrological forecasting with a hydrological analogues method (e.g. Svensson, 2016), and the model set-up is already being applied in seasonal forecasting using an Ensemble Streamflow Prediction approach in the UK Hydrological Outlooks (www.hydoutuk.net). This will also be added to the manuscript.

Parry, S., Wilby, R., Prudhomme, C., Wood, P., McKenzie, A. (2018) Demonstrating the utility of a drought termination framework: prospects for groundwater level recovery in England and Wales in 2018 or beyond. Environmental Research Letters.

Lavers, D., Hannah, D., Bradley, C., (2015) Connecting large-scale atmospheric circulation, river flow and groundwater levels in a chalk catchment in southern England. Journal of Hydrology 523, 179-189. Svensson, C. (2016) Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues. Hydrological Sciences Journal. 61 (1), 19-35.

Most importantly, your method relies on the fact that a large volume of high quality meteorological observations (for both P and PET) were available in the early 20th century, before river discharges were systematically gauged. This was definitely the case for the UK, but in order to argue that the approach you propose is applicable beyond the UK (which would make it more relevant to the global hydrologic community), you have to show that what happened in the UK is not an exception. It can very well be that met data was collected way before flow data in other countries as well, but you have to make this argument explicit (and ideally back it up with some data).

RESPONSE: We believe that it is common that met data records begin before hydrological data records (within Europe at least), simply due to the relative complexities of recording temperature and rainfall over river levels or flows. Newly digitised observed climate datasets (such as the one employed in this study) are becoming increasingly extending observed series held by met services across Europe. Furthermore, Caillouet et al (2017) made use of modelled climate reanalysis data, and the approach could also be applied to other long term reconstructed climate datasets (such as the monthly Casty et al 2007 data). This comment will be added to the manuscript.

Caillouet, L., Vidal, J. P., Sauquet, E., Devers, A., and Graff, B.: Ensemble reconstruction of spatio-temporal extreme low-flow events in France since 1871, Hydrol. Earth Syst. Sci., 21, 2923-2951, 10.5194/hess-21-2923- 2017, 2017.

Casty, C., Raible, C. C., Stocker, T. F., Wanner, H., Luterbacher, J.: A European pattern climatology 1766-2000: Climate Dynamics, 29, 7-8, 791-805, 10.1007/s00382-007-0257-6, 2007.

Minor comments

L210 I am not sure I understand your multi-objective approach to select catchments. How do you weigh different criteria when ranking the parametrization (e.g., how do you differentiate a parametrization A with a NSE of 0.64 and a Q95APE of 34 from a parametrization B with a NSE of 0.70 and a Q95 APE of 40 – which one dominates?). What optimality concept is your approach consistent with (pareto, maxi-min (i..e maximizing the worst performing metrics), . . .)

RESPONSE: The ranking was done as simply as possible, and does not conform to a traditional optimality concept due to the need to rank by 6 metrics at once.

The matrix of 500,000 parameter sets and their scores was sorted first by NSE and a rank column was added giving each parameterisation a rank (1 best to 500,000 worst); the matrix was then sorted again but by logNSE and a new rank column was added; then again by absPBIAS etc. until there were 6 rank columns, one for each metric. The ranks were then summed, and the matrix was ordered by this total rank (with the lowest number being the best parameter set).

However, we found that this left us with a sub-optimal scoring system, as slight improvements in one metric were occasionally outweighing more severe degradations in other metrics, e.g. absPBIAS scores better by 0.001 but NSE scores worse by 0.1). This is why we then set the thresholds. We took the ranked matrix, and starting at the top, looked down the rows of parameterisations until we found one that met the hardest threshold criteria for all 6 metrics. If this was not the originally top ranking parameterisation, it was bumped to the top of the list, and the search was run again. If a second parameterisation was found to meet all 6 criteria, it was then bumped to second place, and the search was run again. Etc.

This created a matrix where all parameterisations that met the hardest criteria were at the top of the list (ordered by their original rankings), followed by those that met the middle criteria (ordered by their original rankings), followed by the softest etc.

This was done for each catchment individually.

As mentioned earlier, we will endeavour to clarify this in the revised manuscript, and will likely provide the R code.

There are lots of acronyms to remember. A Table summarizing the abbreviations would be useful

RESPONSE: We will consult with the editors and include a table of acronyms in

the supplementary information, if appropriate for the journal.

Fig 5: labelled pointers showing the catchment that you specifically discuss in the text would be useful.

RESPONSE: We will add markers to the figure

L132, 502: Please refrain from citing work in preparation.

RESPONSE: We will remove these references.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-3, 2019.