

Reply on RC2

Jamie Hannaford et al.

Author comment on "The enhanced future Flows and Groundwater dataset: development and evaluation of nationally consistent hydrological projections based on UKCP18" by Jamie Hannaford et al., Earth Syst. Sci. Data Discuss.,
<https://doi.org/10.5194/essd-2022-40-AC2>, 2022

We thank the reviewer for these helpful and constructive comments which will guide the revised manuscript. We reply to these comments in turn below.

This manuscript deals with a dataset of multimodel hydrological projections across the UK. It presents in a detailed way how this dataset has been produced, but it lacks a description of the dataset itself. I would therefore recommend a major revision for the authors to bring in an overview of what's actually in this (otherwise important) dataset. I would therefore also appreciate more detailed comments on how to use it, notably for interpreting hydrological projections where models have been calibrated against anthropogenically disturbed streamflow observations.

>>> We describe the dataset in section 9. For how to use the data, this is the subject of a series of separate demonstrators we describe briefly, but which are being written up elsewhere. For this data paper we cannot highlight all use cases, and we have already highlighted some of the uses of the original FFGWL data. We agree that we could add a few short paras highlighting uses of the data and appropriate caveats. We will clarify the issue with anthropogenic disturbances, as highlighted by reviewer 1 too.

Main comments

- *Calibration period:* Despite (too) long descriptions of hydrological model set-ups, there are important missing information on the calibration of conceptual models. First, I haven't seen the calibration period explicitly mentioned in the text. I assume that is the complete historical 1961-2018 period covered by simobs simulations, but this is to be made explicit.

>> also noted in response to R1, we will clarify.

- *Parameter transferability:* the previous comment calls for another question on the transferability of parameters of conceptual models, which has been shown for some time as an important issue when dealing with climate change (see e.g. Thirel et al., 2015). This issue is unfortunately completely absent from the manuscript, even in the discussion part. This issue is often dealt with by using more or less advanced split-

sample set-ups, but many alternative propositions have been made over the recent years (e.g. Todorović et al., 2022). This issue should therefore be dealt with in the manuscript, at the very least as a comment in the discussion part.

>>> we will clarify in the method and also in the discussion, and will add the references highlighted as good examples, thank you.

- *Calibration on catchments with anthropogenic disturbances:* This issue is seemingly considered as a rather light one in the manuscript (L604-615). I rather disagree here, as the main underlying hypothesis is not even mentioned here: models calibrated on influenced data will deliver hydrological projections of influenced streamflow in which the amount and seasonality of anthropogenic disturbances (abstractions, reservoir management and so on) is equal to those during the calibration period. Which is clearly an unrealistic feature of the future. This makes the understanding and the use of such projections quite difficult for water managers and stakeholders. I hope that communication around the eFLaG dataset can handle this issue, but it is definitely not convincing in the manuscript.

>>> we will clarify in the method and also in the discussion. Of course, we recognise that disturbances will change in future. But it is not really our intention to accurately model catchment flows to the 2080s – rather, to look at potential climate futures for each catchment given current conditions, for planning purposes, and as widely adopted in many other applications (including the original FFGWL, the model for eFLaG). We will modify this accordingly in the intro and the discussion, as also noted in our replies to R1.

- *Description of dataset:* as already pointed out by Anonymous Referee #1, this manuscript describes how the eFLaG dataset has been built (which is definitely commendable), but it does not describe the dataset itself. Indeed, there is no e.g. (1) overall summary statistics (temporal or spatial) of present-day-period simulated streamflows or groundwater levels, (2) overall summary statistics of projected evolution or changes over the UK, and (3) no case study example of the numerous time series produced. All these features are essential to get a grasp of what's in the dataset and are therefore in my view a required feature of a data paper.
- *Length of model descriptions:* I guess that the manuscript is currently quite lengthy because of the extent to which hydrological model set-ups are described, at the expense of the more general and important issues listed above. A new balance should be reached in the revised version of the manuscript.

>>>We feel it is more important for users of the data paper to understand the provenance (i.e. methodology) rather than the outcomes, which are the subject of multiple papers and reports.

Specific comments

- Table 1: I disagree with the partitioning of uncertainties here: “model structure” or “model choice” (or here “hydrological models” like “climate models”) are equivalent. GR4J and GR6J are indeed different models, probably a bit similar to each other than to PDM for example, but “model structure uncertainty” is commonly used as opposed to “model parameter uncertainty” in common uncertainty decomposition of

hydrological projections (see e.g. Christierson et al., 2012 for the UK). This relates to one of my main comment on parameter temporal transferability.

>We are following the logic and nomenclature of the Smith et al. 2018 paper here. We see GR46J and GR6J as two structures. We will add a line on parametric uncertainty but note we do not sample it (see replies to R1) – this was on an earlier version but omitted.

- L178-189: Please recall a reference for UKCP18

>>will do, we refer to Murphy et al. at line 55 but will reiterate here

- L190-196: This issue with Hadley Centre models has been around for the last 15 years at least... But here it means that simrcm streamflow series also have 360 days per year? This is what I can see from the data files, but this is not discussed or even mentioned in the manuscript. I wonder what a water manager would say when looking at those files... It would therefore be necessary to rise the issue in the manuscript and also provide some advice for water managers and stakeholders on how to use such unusual time series, in order to prevent any misuse of even rejection based on lack of credibility.

>>>We will highlight this in the text, and also in the discussion section (as noted with R1 replies, we will add a short 'applications' section. The issue of 360 day years is also highlighted in the demonstrators, with appropriate recommendations.

- L215-217: This spatial disaggregation step is not clear enough. Plus, what is the standard-period in SAAR? And why is HadUK-Grid not used here? All these choices are not enough commented and justified.

>> This is a very standard approach but we will add extra text to clarify this, as detailed in replies to R1. SAAR is used as a standard, not subject to change as HadUK is.

- L228-236: Could you give here a simple description of the PET formula (e.g. "Penman-Monteith-like")? Indeed, the choice of PET formula may have strong consequences especially for low-flow changes (see e.g. Lemaitre-Basset et al., 2022). This choice would also deserve a comment in the manuscript.

>> This follows the CHESS Method as highlighted with a reference so we did not expand, but we will add a short additional sentence to make clear. We agree PE formulation can be important and we highlight this already in Section 8, L849. Thanks for the additional reference.

- Figure 2: The bias-correction factors are quite high for some month/model. This should also deserve a comment, especially with the somewhat overlapping issues of model weighting versus internal variability.

>>> we will add a comment about the high bias correction factors

L291-294: In relation to one of my main comments, I could not find in the eFLaG_Station_Metadata.xlsx file any flag indicating a near-natural catchment (e.g. belonging to UKBN2) or borehole that would help identifying locations where streamflow/groundwater projections are natural streamflow/groundwater projections. This lack of flag (I noted the FARL field, but this is far from being the only relevant source of disturbances) makes me uncomfortable with this definitely rather non-homogeneous dataset.

>>>We do highlight UKBN2 membership in the spreadsheet (Column I). We do not refer

to other sources such as NRFA descriptions or FAR (Factors affecting runoff) codes, but we will add a sentence to note these sources are available.

- Figure 3: The text is very small and makes maps difficult to read.

>> we would hope this would be made a large map in the paper. We cannot really make the text much bigger without making it too cluttered.

- L395: The CHESS version cited here has been superseded.

>>we will highlight this

- Table 3: It is necessary to have this table in the main text?

>>> yes, it is necessary to allow the reader to look at the metrics close to where they are cited. This would be our preference.

- L555-559: How is SGI used for evaluation? This is unclear until Figure 5 a few pages later when we learn about the NSE_SGI.

>>>SGI is used as the basis for comparison of the observed and simulated data, with the various metrics used to establish performance. We will make this clearer.

- L634: already written above.

>>not sure what is meant here, there is a general point made and then a specific one wrt Rudd et al. 2017.

- Figure 5, caption: "NSE_SGI"

>>agreed we need to make the caption and text consistent

- Figure 7: This figure is definitely too small for it to be correctly read and interpreted (see e.g. L698-700). What about using instead a metrics (or a very few metrics) based on the correspondence between FDCs? This would allow showing all locations in the main text.

>>we prefer to show for individual catchments. We imagine this would be a large image in the paper (see also appendices).We do in fact show metrics based on correspondence of specific FDC quantiles, e.g. Figure 9 for Q90 and then the various equivalents in the Supplementary info.

- Figure 9: This is a poor choice of color scale which makes contrasts much too difficult to read. I would highly recommend using one of the scales available at <https://colorbrewer2.org> and recommended by the IPCC (2018), and reserve high values for darker colors, for them to be emphasized. Plus, the legend is repeated.

>>>We agree this could be revised, thanks for the suggestion. We can use a palette from colorbrewer2, but to be honest, it'll still be difficult to distinguish between adjacent categories -- that's the nature of these palettes, and we feel that it does not effect the interpretation. Re: Legends, there are two legends that are almost identical but one is for Q90 (for the 4 HMs) and the other is for L90 (for g/w)

- Figure 10: Same comments as above. Impossible to distinguish between 2-5, 5-10, and

10-20 classes.

Technical corrections

- L105: "EdgE"
- L126: "a,b,c"
- Table 1: Please define "PPE" (Perturbed Physics Ensemble)
- L274: Please define "EIDC"
- L644: missing "models"
- L694: "re"?

>>> thanks, we will make these changes