

Reviewer 1

The authors have done a good job at responding to points raised in the first review. In particular they have dealt with points of clarification and added further discussion on the limitations of the dataset, especially regarding the transferability of hydro model parameters. Overall it is my conclusion that the paper is ready for publication with some further minor corrections, which I outline below. I am happy for the editor to sign off on these.

>>we thank the reviewer for this assessment of the improved manuscript

In my first review I noted that the title, abstract and intro of the paper gives the sense to the reader that the paper will present the final dataset of projected change in flows, which the paper does not actually do. Rather the paper describes the development and evaluation of the dataset while further papers are in train to describe the final data and projected changes. The authors have edited the title in this review but I still think that this does not actually say what the paper does. This could be easily addressed as follows:

Change the title to: The eFLaG dataset: development and evaluation of nationally consistent projections of future flows and groundwater based on UKCP18

>>We agree that this helps so we have changed to this

Abstract: The abstract begins with 'this paper presents the dataset'. It doesn't, it presents the development and evaluation of the methods to produce the dataset. It is an important difference. The abstract should therefore open with 'This paper details the development and evaluation of...'

>>We have changed this too

Introduction: same point as abstract, need to emphasise development and evaluation in the opening sentence.

>>As above

The biggest limitation of the dataset is use of only a single climate model and emission scenario. This is certain to underestimate the range of plausible change in future flows. While these points are dealt with in the limitations section they are important, especially given the intention of informing water management. Therefore I suggest that a sentence on key limitations is included in the abstract.

>>We have added that to the abstract

At bottom of page 6 (I am reading the tracked changes version) it is stated that model structure uncertainty is accounted for by considering two versions of one of the models. This is a minor part of this evaluation, inclusion of physically based gridded model and PDM is more structurally different than two versions of the GR models. This sentence needs to be edited to simply saying uncertainty in model structure is accounted for.

>>We considered the difference between GR/PDM and G2G to be model uncertainty (diff models) and ArGR to be structural (in the sense of different structures of same model type). But agreed there are not always clear distinctions between this. We have edited accordingly.

Page 8 when describing biases in precipitation, give some numbers in the text to help reader quantify the magnitude of biases, at present only descriptives are provided such as substantially over estimates.

>>we have added some context

The authors note that the approach taken in developing the study design vis uncertainties was to crystallise the uncertainties. Going back to the Smith et al paper crystallise is described as sampling the spectrum of potential results at each step to reduce or "crystallize" salient outcomes. The selection of a

single climate model/emissions scenario and then more heavily sampling hydrological model components of the cascade does not seem to me to match this description.

>>We agree this is technically not correct, and amended accordingly

Page 16 reduce the number of times the term described below is used, at least three times in quick succession.

>>we have changed this

Page 17 reduce the number of times the term a range of metrics is used, at least three times in quick succession

>>we have changed this

Page 22 What are increasing aggressive stages?

The stages of the calibration optimise increasing numbers of model parameters in turn (which is what was meant by increasingly aggressive). The text (repeated below) has been updated to make this more explicit. Full details are included in the Supplementary info.

"Parameter estimation was performed using an automatic calibration procedure that applied a simplex optimisation scheme (Nelder and Mead, 1965) to increasing numbers of model parameters in turn."

Reviewer 2

80: See Chan et al for how climate projections have been used in drought and climate change assessments: <https://doi.org/10.1177/03091333221079201>

>>We added this ref

128: However, they have been used in industry in support of WRMP24

>> we meant in terms of the scientific literature, but we can highlight this – we referred to Thames water's WRMP. We note that these are draft and still not readily available.

328: Sites or/and catchments?

>>We use interchangeably at various points, we think this is clear

333/417/943: Artificial influences captured indirectly via calibration on observed flows (make this clearer from the beginning). Standard industry approach is to do climate change assessment on naturalised and then denaturalise using artificial influences.

>> We thought we did make this clear right at the start of the modelling section now (417), and we do ref it indirectly at 333 when talking about catchments. Hard to see how we could ref earlier without being 'too early' before the context.

515: Descriptions of models are quite long and could be moved to supplementary material

>>we disagree, there is a lot of SI and we felt the need to keep a minimum of the model description in the body. We trimmed superfluous material last time. We feel this should now stand (reviewer 1 picked this up last time but not again this time).

590: Where was the spring factor used – were there 'rules' used to identify relevant catchments?

In PDM, the spring factor is a multiplicative factor on the baseflow used to represent net losses or gains of water to permeable catchments (e.g. from external springs). Because of this we used a criteria based on the Base Flow Index (a measure of the relative proportion of base flow), $BFI > 0.7$, to identify potential catchments for which the use of the spring factor might be hydrologically plausible (the only exception being the Leven at Linnbrane, strongly affected by Loch Lomond, where we deemed its use inappropriate). Then, for such catchments, we chose either to use a model configuration using the spring factor, or one using the rainfall factor, according to which produced the best performance in the calibrated modelled flows. Full details are included in the Supplementary info.

983: add apostrophe in "model's"

>>> Done

984: this is particularly relevant in future with changing rainfall regimes

>>we will add a sentence to this effect.