Response to reviewers

General comments

This paper reports on the development of a generic algorithm for water supply dams and its application to a hydrological model for the United Kingdom. The development of hydrological models that include reservoir operations is essential for estimating flow rates and water resources. The paper argues that a relatively simple model can represent the operation of water supply dams in the UK. The results reported are generally considered reasonable, but some of the descriptions were difficult to understand and should be refined.

Thank you for your review and helpful comments, we will address each of them in the response below.

Specific comments

Line 185, ”three equal classes of slope and accumulated area”: This part is hard to read. What do you mean by “equal classes”? What is “accumulated area”?

Thanks, we have clarified this in the manuscript as follows:

In this study, HRUs were classified using a 2.2-km input grid (consistent with national climate projection data), which were further sub-divided by gauged sub-catchments (which include those defined by reservoir nodes) and percentiles of slope and upslope accumulated area (i.e. the area of land draining to a particular point in the landscape). This ensures that HRU’s cascade downslope to the bottom of the valley and makes sure that the spatial variability of the climatic inputs is appropriately represented.

Line 218, “to determine unbiased values for the natural model parameters in reservoir catchment”: What do you mean by “unbiased”?

By using the term ‘unbiased’ we refer to the fact that DECIPHER’s seven standard model parameters (which have been designed to simulate the hillslope hydrology and are referred to as ‘natural’ model parameters) should not be overcompensating for any reservoir processes. We agree that the term ‘unbiased’ may be confusing so we have removed it such that the sentence now reads:

The top 10 natural transfer function parameter combinations were then chosen by calculating the non-parametric KGE (Pool et al., 2018) (see section 3.6) in all near-natural catchments. The 10 combinations with the highest average non-parametric KGE across all the near-natural catchments were subsequently used to determine the natural model parameters in reservoir catchments.

Line 234, “the OS river layer”: What is this?

We have clarified this in the manuscript:

We use a 50-m gridded digital elevation model (Intermap Technologies, 2009) to generate the river network in DECIPHeR, extracting headwater cells from the an open-access river network which maps the rivers across GB generated by the Ordnance Survey (Ordnance Survey, 2023).

Line 252, “5000 times in the reservoir scenario sampling reservoir parameters”: I couldn’t understand these runs and how they were implemented. What do you mean by “sampling”?

‘Sampling’ refers to the random selection of the 5000 parameter combinations chosen from within the set bounds. We have altered the text here to make this clearer:

The no-reservoir scenario runs 10 simulations in each reservoir catchment using the top 10 natural transfer function parameter combinations (section 3.3). In the reservoir scenario, for each of the same 10 parameter combinations, we sample the reservoir parameters 500 times, resulting in 5000 simulations per catchment.

Line 264, "In this study abstraction and compensation flow remain constant throughout the simulation." This is an interesting point. It implies that the summation of the abstraction and the compensation flow must be equal to the mean annual inflow (otherwise, the long-term water balance does not close). The key difference between this study and Hanasaki et al. (2006) is whether to separate the outflow of reservoirs into abstraction and compensation flow. This point would be mentioned somewhere.

We agree that the one key difference between our operating rules and the Hanasaki operating rules is that our rules include a public water abstraction which is removed directly from the reservoir (and therefore the water balance is not closed within the catchment but is exported for external water supply). This is a very important component of the reservoir operations in GB. We have distinguished this important difference between our rules and the Hanasaki rules in a new paragraph which has been added to the Model Evaluation section (to avoid repetition, see our response to your final comment for a copy of this paragraph).

Line 304, "The results from the near-natural simulations for the best nationally-consistent set of transfer function parameters": It is a bit awkward to read this. Why don't you put a simulation name? I think the combination must be [NATural or REServoir], [INDividual or national-CONsisnt]; hence, this run might be called "NAT-CON."

Thank you for this suggestion but we would prefer to avoid introducing more acronyms.

Line 315 "(or simulations, decided based on those with the highest median non-parametric KGE across all 137 catchments)": Hard to read. Better to clearly explain/define this in the Methods section.

We have added the following to the end of section 3.3 which describes these runs in the methods section to make this clearer:

The top 10 natural transfer function parameter combinations were then chosen by calculating the non-parametric KGE (Pool et al., 2018) (see section 3.6) in all near-natural catchments. The 10 combinations with the highest average non-parametric KGE across all the near-natural catchments were subsequently used to determine the natural model parameters in reservoir catchments.

Line 333, "to define the transfer functions used in this study": It is a bit confusing. I believe at least two transfer functions are used in this study: one for hydrological parameters and the other for reservoirs. For better readability, these should be clearly distinguished.

We agree this is difficult to distinguish. We have now altered the manuscript to specify whether we are referring to the natural transfer functions or reservoir transfer functions.

Figure 3: Please confirm whether the label "Top national simulation" is correct. I expected "Best national simulation". Anyway, the mixture of top and best is confusing for me.

Thank you for pointing this out. We will be more consistent with our language here and use only 'Top'.

Line 338-339 (Equations 6-7): Again, this is a very interesting point. The compensation flow is a function of the catchment area. Because the inflow to a reservoir is basically proportional to the catchment area, it can be said that this equation is similar to Hanasaki et al. (2006; CF ~ mean annual inflow). The abstraction is a function of reservoir capacity. Because the reservoir capacity is only weakly correlated with the inflow usually, it can be said that this formulation is different from Hanasaki et al. (2006). Then, let me come back to my previous point. The equations 6-7 do not seem to guarantee the water balance shown in Equation 1. From this perspective, it is doubtful that the overall modeling framework is hydrologically reasonable. This point should be additionally discussed in the manuscript.

The water balance shown in equation 1 is guaranteed by equations 4 and 5 which ensure that the abstraction and compensation flows are reduced to zero if the storage is too low. We agree that the CF should be proportional to the mean annual inflow and we also agree that by including an abstraction flux our rules are different from those introduced by Hanasaki. We will make this clearer by distinguishing the key differences between our rules and the Hanasaki rules in a new paragraph which has been added to the Model Evaluation section (to avoid repetition, see our response to your final comment for a copy of this paragraph).

Line 341 “is in line with the observed data”: What does it mean by “in line with”? The gray dotted line is basically above the blue dots.

To check that our natural and reservoir functions were generating physically realistic parameter values for the CF and abstraction parameters (and to define the transfer functions in the first place) we compared the parameters identified by our transfer functions to the CF and abstraction values recorded in the literature (where this information was available). By saying that our chosen transfer functions identify ABS parameters “in line with the observed data” we mean that the parameters chosen by the best nationally-consistent transfer functions are similar to (or in line with) those recorded in the literature.

That said, this is only true for the ABS parameter and we agree that the gray dotted line is above the blue dots. We found that the model was very insensitive to the choice of CF parameter and so although the best nationally-consistent transfer functions identify ABS parameters which are in line with the observed data the same is not true for the CF and we will point this out here.

We have rephrased this section of text below to account for your following two comments as well as this.

Line 343 “at the upper end of the sampling limits”: What are the “sampling limits?”

These are introduced in Table 1. These are the bounds within which the transfer function parameters are sampled. We now reference this table in the text which we have rephrased below.

Line 343 “the range of variability of the transfer functions associated with the top 5% of national-consistent simulations”: Top 5% in terms of what? What do you mean by “the range of variability”?

Here we are referring to the 5% of simulations with the highest average KGE value (across the whole sample of catchments). The range of variability refers to the amount of variability in the shape of the transfer functions within this top 5%. We will make this clearer in the text and agree it could be much better described.

All three of the previous comments have been addressed by being rephrasing the text to read:

The top nationally-consistent calibration associated with the ABS parameter (marked on Figure 4a with a grey dashed line) generates parameters which are similar to those observed in the literature. However, the top nationally-consistent transfer function selected for estimating the CF parameter (marked on Figure 4b with a grey dashed line) lies close to the upper end of the sampling limits (Table 1) and does not match the observations. To investigate the sensitivity of the model to each of the reservoir parameters (CF and ABS) Figure 4 also shows the variability in the transfer functions associated with the top 5% of nationally-consistent simulations (this is displayed on Figure 4 with darker shading). The top 5% of simulations are those with the highest average non-parametric KGE (calculated across the full sample of reservoir catchments). This shows that the model's predictive performance is more sensitive to ABS (p1) than CF (p2).

Line 376 (Figure 4 caption) “top reservoir simulation”: Again, what is the difference between “top” and “best”?

As above, we will be more consistent with our language here and use ‘top’.

Line 502- “the simplicity of our operating rules”: Ideally, the author should also demonstrate that their model outperforms the even simpler model (Hanasaki et al. (2006), which requires no calibration parameter).

We have now demonstrated for four key reservoir catchments the differences in model simulations between our reservoir rules and Hanasaki’s rules. This is demonstrated by two figures and a table (see below) which have been added to section 8 of the supplementary information. Our results reveal some inherent problems associated with using the Hanasaki rules to simulate flow downstream of water supply reservoirs and show that in three of the four examples the Hanasaki rules are outperformed by not just our operating rules but also the model with no reservoir representation. We have added the following paragraph to the model evaluation section to highlight this new analysis:

For completeness, in a few selected catchments we also compared our operating rules to the widely used non-irrigation reservoir rule introduced by Hanasaki et al. (2006). However, since the Hanasaki rule assumes that no abstractions are taken directly from the reservoir, these rules are not well suited to water supply reservoirs in our domain (see section 4.3 and section 8 in the Supplementary Material). We could not compare our operating rules to any of the data-driven approaches in the literature (e.g. Turner et al. (2020)) since their high data requirements could not be fulfilled at the national scale in GB. Comparing the simulations which use our new operating rules to the simulations of the pre-existing hydrological model without reservoir representation thus remains the most feasible and relevant way to evaluate the new proposed model.

We have also added the following paragraph to section 4.4.2 which references these findings:

Finally, Section 8 of the Supplementary Material reports a comparison of the top nationally-consistent simulation with the widely used Hanasaki rule (Hanasaki et al. 2006) in this selection of catchments. Although the Hanasaki rule has no calibrated parameters and is therefore arguably simpler than ours, we found that it delivers a much poorer performance. This is largely because the Hanasaki rule does not allow for abstraction from the reservoirs which is a key component of reservoir operation in most GB reservoirs.

The following has been added to the supplementary material:

To compare the new water supply reservoir operating rules introduced in this paper to an alternative set of rules, Figures S12 and S13 show the hydrographs associated with simulations using both our *new rules and the rules defined by Hanasaki et al. (2006). We find that the Hanasaki rules are not well suited to simulating flow downstream of water supply reservoirs, largely because they have no abstraction component which is a key part of operations at water supply reservoirs in GB. By not accounting for the abstractions taken from the reservoir, reservoirs simulated by the Hanasaki rules are full much more often and therefore often spill in periods where in reality only the compensation flow is released. Table S3 compares the non-parametric KGE scores associated with both the Hanasaki rules and the rules introduced in this paper. It is clear from this table that the Hanasaki rules are not able to well represent the water balance and at three of the four featured gauges where the non-parametric KGE achieved with the Hanasaki rules is lower than what is achieved by a model with no reservoir representation.*

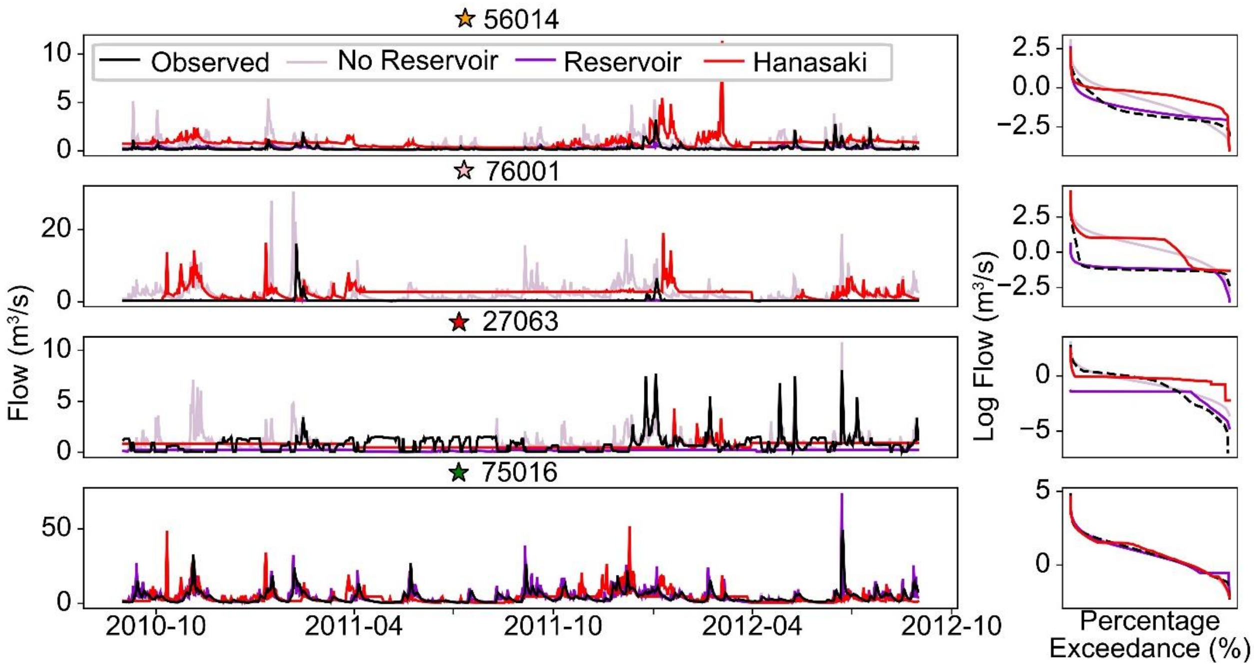

Figure S12. *Hydrographs and flow duration curves top median simulation (nationally-consistent calibration) for selected reservoir catchments compared to simulations using the Hanasaki non-irrigation reservoir rules (Hanasaki et al. 2006).*

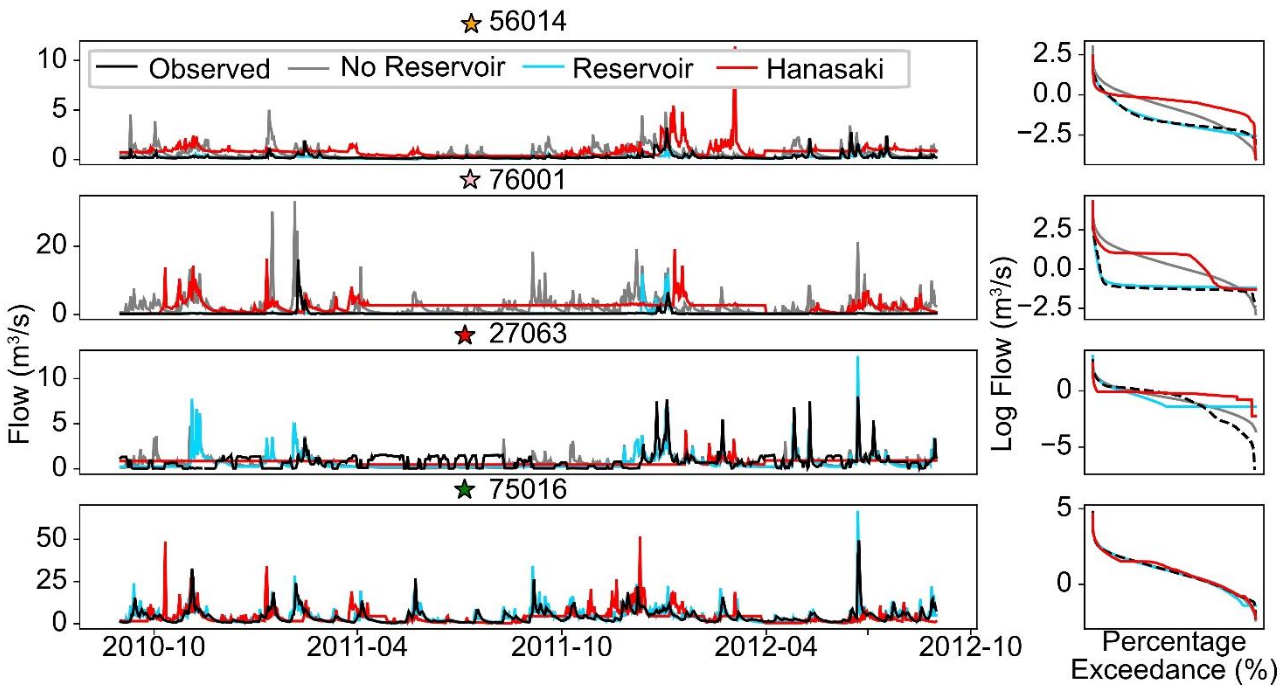

Figure S13. *Hydrographs and flow duration curves from the top individual simulations (catchment-by-catchment) for selected reservoir catchments compared to simulations using the Hanasaki non-irrigation reservoir rules (Hanasaki et al. 2006).*

Table S3. *Non-parametric KGE and associated components for simulations at four featured gauges using water supply operating rules (with both catchment-by-catchment (CBC) calibration and nationally-consistent (NC) calibration), no reservoir representation and Hanasaki et al. (2006) non-irrigation operating rules.*

Catchment	Operating rules	Water balance	FDC	Spearman's rank	Non-parametric KGE
56014	Water supply rules (CBC)	1.00	0.96	0.69	0.69
	Water supply rules (NC)	0.85	0.83	0.68	0.60
	No reservoir representation (NC)	1.93	0.89	0.67	0.01
	Hanasaki non-irrigation rules	1.93	0.69	0.40	-0.15
76001	Water supply rules (CBC)	1.04	0.92	0.45	0.44
	Water supply rules (NC)	0.63	0.62	0.43	0.23
	No reservoir representation (NC)	4.49	0.68	0.44	-2.54
	Hanasaki non-irrigation rules	4.49	0.63	0.35	-2.61
27063	Water supply rules (CBC)	0.83	0.80	-0.08	-0.11
	Water supply rules (NC)	0.24	0.70	-0.16	-0.42
	No reservoir representation (NC)	0.98	0.89	-0.17	-0.17
	Hanasaki non-irrigation rules	0.98	0.73	0.11	0.06
75016	Water supply rules (CBC)	1.01	0.97	0.89	0.88
	Water supply rules (NC)	1.01	0.94	0.87	0.86
	No reservoir representation (NC)	1.04	0.95	0.87	0.85
	Hanasaki non-irrigation rules	1.04	0.93	0.83	0.81

Review Comments:

This paper describes a simplified reservoir operating scheme that is distinctly oriented toward smaller non-irrigation reservoirs that only uses two parameters. This reservoir scheme uses general catchment and reservoir parameters that are more readily available than reservoir time series data. To test the sensitivity and performance of this reservoir scheme, the authors derived parameters for water supply reservoirs throughout Great Britain that encompass reservoirs of multiple sizes and uses. They also analyzed the feasibility of using transfer functions to create a nationally consistent parameterization. In their work, they evaluate the impact of two calibration methods one within each catchment and the other across the entirety of Great Britain. Results show that the methodology works well in catchments with primarily water supply uses and is limited in catchments with multi-purpose reservoirs. This article does a really good job of creating simplified curves and evaluating metrics for evaluating reservoir dynamics with a unique focus on water supply reservoirs.

Thank you for your useful review, we will address each of your comments individually below.

Major Edits:

Equation 4 and 5 have Smin, which I assume refer to dead storage, however, this variable is not defined in the text. Please clarify this in the text so the reader is better able to follow what the equations are referring to.

We have now defined this in the text, thank you.

Section 3.5.1: As readers may be unfamiliar with the study domain, it could be useful to have a figure that shows the general makeup of the reservoirs used in this study colored by use, size and a second panel with the naturalized and non-naturalized catchments and general characteristics. This would allow a reader who is unfamiliar with the catchments in GB have a better idea of the characteristics of the catchments mentioned in the paper.

Thank you for pointing this out. We have now added the figure below to section 3 which shows the size and distribution of water supply reservoirs across GB and the reservoir and natural/ near-natural catchments.

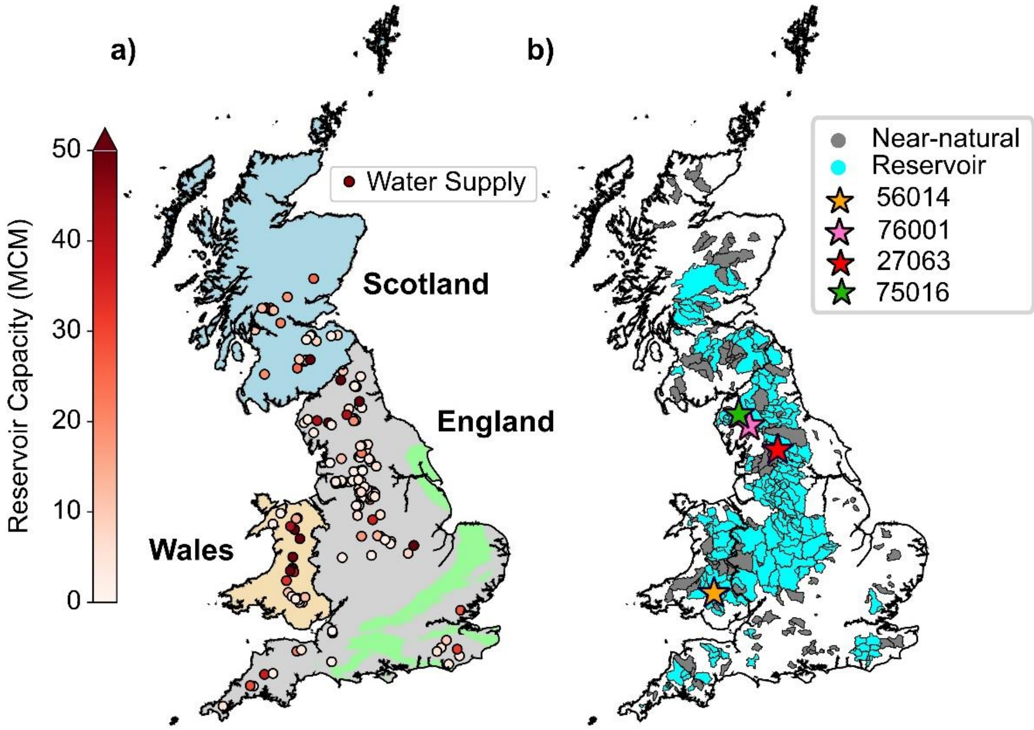

Figure 2. *Distribution of (a) water supply reservoirs across GB and (b) near-natural and reservoir catchments used in this study. Reservoirs are coloured by their storage capacity and the four catchments featured in Figures 6 and 7 are highlighted with stars on subplot b.*

Line 308: I am personally not familiar with the catchments in GB. To increase clarity, I would include a catchment map either with numbers or highlight these catchments in Figure 2 so that the reader knows which area you are referring to. Alternatively, you could leave out the reference to these specific catchments and describe them more generally. The same also goes for Line 388, 390, and 460. I would make the distinction as to why you picked these catchments and give an overview of them prior to mentioning them so the reader is not confused.

Thank you for this suggestion. We now describe these catchments more generally as you have suggested, in this instance we agree that the catchment characteristics are more important than their specific location. For example, at line 308 the text has been altered to read:

For example, the Aldbourne at Ramsbury (39101) and the Ewelme at Ewelme Brook (39065) (which are both chalk catchments)

We will also make sure that all references to the four featured catchments (denoted by stars on Figures 4-7) are highlighted in the text and we have added the locations of these catchments to the new Figure 2 introduced above. For example, on line 388 where we reference the Haweswater Beck at Burnbanks (76001) we will add (denoted by a pink star on Figures 2 and 5-8).

Line 313: This is an interesting note about reservoir construction in GB and it's impacts on your work. What would be the larger impacts on other regions that are more groundwater regions dominated such as the southwestern US or for dams built on limestone or more porous rock such as the Mosul Dam in Iraq?

We are currently coupling the surface water components of DECIPHeR to a groundwater model (Rahman et al, 2023) so it would be important to include this model development if the model was applied to reservoirs in catchments with significant amounts of groundwater flow.

Line 337: The rational for using a linear relationship is not clear from the text. Why was a linear relationship better than a non-linear one?

Thank you for pointing this out, have altered the text to read:

Since the observations (Figure 4) do not show any evidence of non-linearity, we chose to use a linear (and hence more parsimonious) relationship for both transfer functions.

Line 350: Including Table 1 with the upper and lower bounds of the transfer function parameters is nice, however, the Table is not referenced at all in the text. Instead of a table, it might be nice to include a figure of the parameter ranges, average parameter value, or another metric to show the regional differences in transfer function parameters across different catchments in GB.

Thanks – we have now added a reference to this table into the text. The upper and lower bounds of the parameter ranges are displayed on Figure 3 but have added the figure below to the supplementary material which shows the regional differences in transfer function parameters by plotting the parameters chosen in each catchment by the catchment-by-catchment calibration spatially across GB.

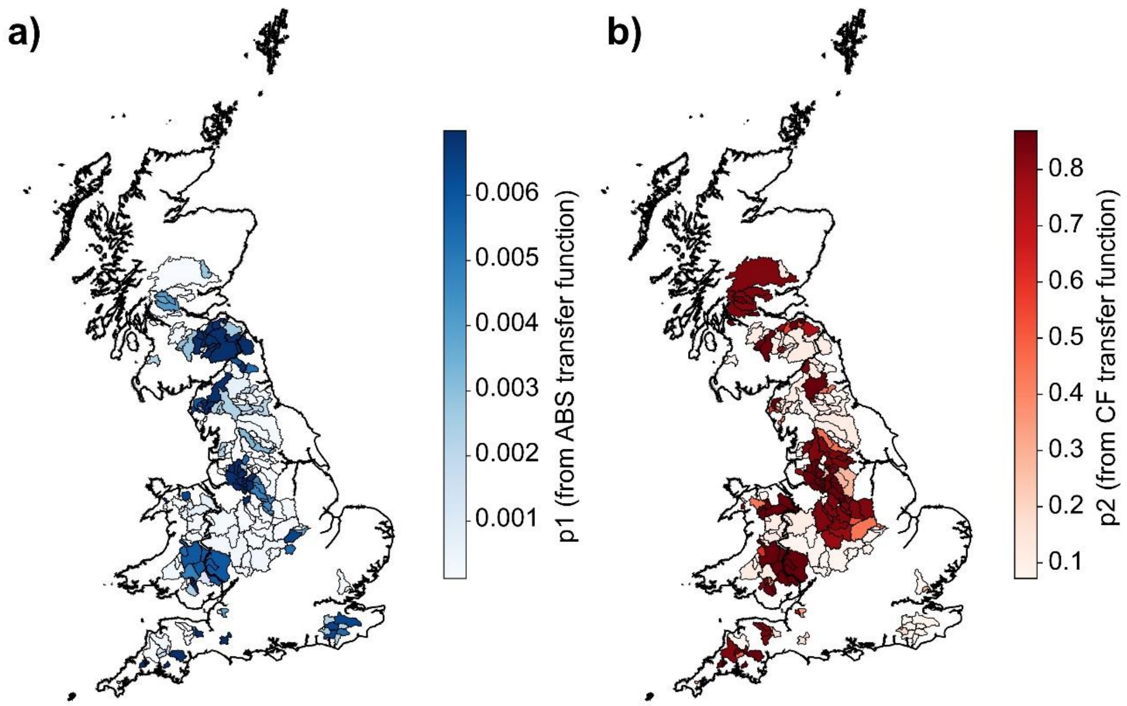

Figure S14. *Spatial distribution of transfer function parameters associated with both CF and ABS. Catchments are colored by the transfer function parameters identified in the catchment-by-catchment calibration.*

Figure 4 and Figure 5: Why did you pick the four catchments denoted by stars in these figures? What characteristics cause those to be chosen for the case studies and do these directly relate to the ones mentioned in Line 308, 388 and 390? If so, I would make the connection between these clearer.

We chose these four catchments because they show the successes and areas for improvement for our operating rules. This description is included in Section 4.4 where these catchments are investigated in more detail, but we have added the following description into the text after the 4 featured catchments and their results are first introduced in Figure 4 to make this clear earlier on:

Figure 5 also highlights four catchments with star markers which are designed to demonstrate where the operating rules are working well and where improvements are needed (see Figure 2 for the location of these catchments across GB). Catchments 76001 and 56014 (pink and yellow stars) show large improvements in the KGE where the operating rules are working well. Comparatively, changes in the KGE at catchments 27063 and 75016 are minimal. This is discussed in more detail in section 4.4.

Only one of the catchments referenced in lines 308, 388 and 390 directly relates to those featured using the stars, but we have pointed this connection out in the manuscript by referencing the star color. Line 388 now reads:

The largest improvement in KGE is 2.99 which is seen at the Haweswater Beck at Burnbanks (76001) (denoted by a pink star on Figures 2 and 5-8) where the metric increases from -2.55 to 0.44, largely driven by the water balance component which decreases from 4.49 to 1.04.

Line 549: Can you expand on what you would suggest other do in cases where your methodology is limited (i.e. in the case of multipurpose reservoirs)? Should a combination of methodologies be used or a more generalized approach such as Hanasaki et al., 2006? Additionally, how would the authors envision upscaling this approach to other hydrologic models?

Thank you for your comment. We discuss this in the limitations and future work section of the discussion but will make this clearer here by rephrasing the text to say:

It is no surprise that our rules do not work well here, where they are likely to miss some crucial coordination with the downstream river and misrepresent the purpose of the reservoir (Rougé et al., 2019). However, future work might consider defining new transfer functions to describe the operating rules at reservoirs in this sample (see section 5.4 for more detail).

Section 4: The authors do a really nice job of explaining their methodology and results. It would be interesting to see a comparison between this simpler model and one or the more generalized approaches or perhaps the approach that is already in DECIPHeR as another point of evaluation.

Thank you for your comment. At present there is no alternative approach for representing reservoirs in DECIPHeR, so by comparing our results to the model with no reservoir representation we are comparing our results to the current approach. We will make this clearer in the manuscript (see text below).

Regarding a comparison to a more generalized approach, this point has been raised by both yourself and the second reviewer so we will add a comparison to the Hanasaki rules into the supporting information (see Figures S12 and 13 and Table S3 below) and add the following paragraph to the Model Evaluation section to address this and highlight the additional analysis:

For completeness, in a few selected catchments we also compared our operating rules to the widely used non-irrigation reservoir rule introduced by Hanasaki et al. (2006). However, since the Hanasaki *rule assumes that no abstractions are taken directly from the reservoir, these rules are not well suited to water supply reservoirs in our domain (see section 4.3 and section 8 in the Supplementary Material). We could not compare our operating rules to any of the data-driven approaches in the literature (e.g. Turner et al. (2020)) since their high data requirements could not be fulfilled at the national scale in GB. Comparing the simulations which use our new operating rules to the simulations of the pre-existing hydrological model without reservoir representation thus remains the most feasible and relevant way to evaluate the new proposed model.*

We now refer to these results in section 4.4.2 with the following text:

Finally, Section 8 of the Supplementary Material reports a comparison of the top nationally-consistent simulation with the widely used Hanasaki rule (Hanasaki et al. 2006) in this selection of catchments. Although the Hanasaki rule has no calibrated parameters and is therefore arguably simpler than ours, we found that it delivers a much poorer performance. This is largely because the Hanasaki rule does not allow for abstraction from the reservoirs which is a key component of reservoir operation in most GB reservoirs.

The following will be added to the supplementary material:

To compare the new water supply reservoir operating rules introduced in this paper to an alternative set of rules, Figures S12 and S13 show the hydrographs associated with simulations using both our new rules and the rules defined by Hanasaki et al. (2006). We find that the Hanasaki rules are not well suited to simulating flow downstream of water supply reservoirs, largely because they have no abstraction component which is a key part of operations at water supply reservoirs in GB. By not accounting for the abstractions taken from the reservoir, reservoirs simulated by the Hanasaki rules are full much more often and therefore often spill in periods where in reality only the compensation flow is released. Table S3 compares the non-parametric KGE scores associated with both the Hanasaki rules and the rules introduced in this paper. It is clear from this table that the Hanasaki rules are not able to well represent the water balance and at three of the four featured gauges where the non-parametric KGE achieved with the Hanasaki rules is lower than what is achieved by a model with no reservoir representation.

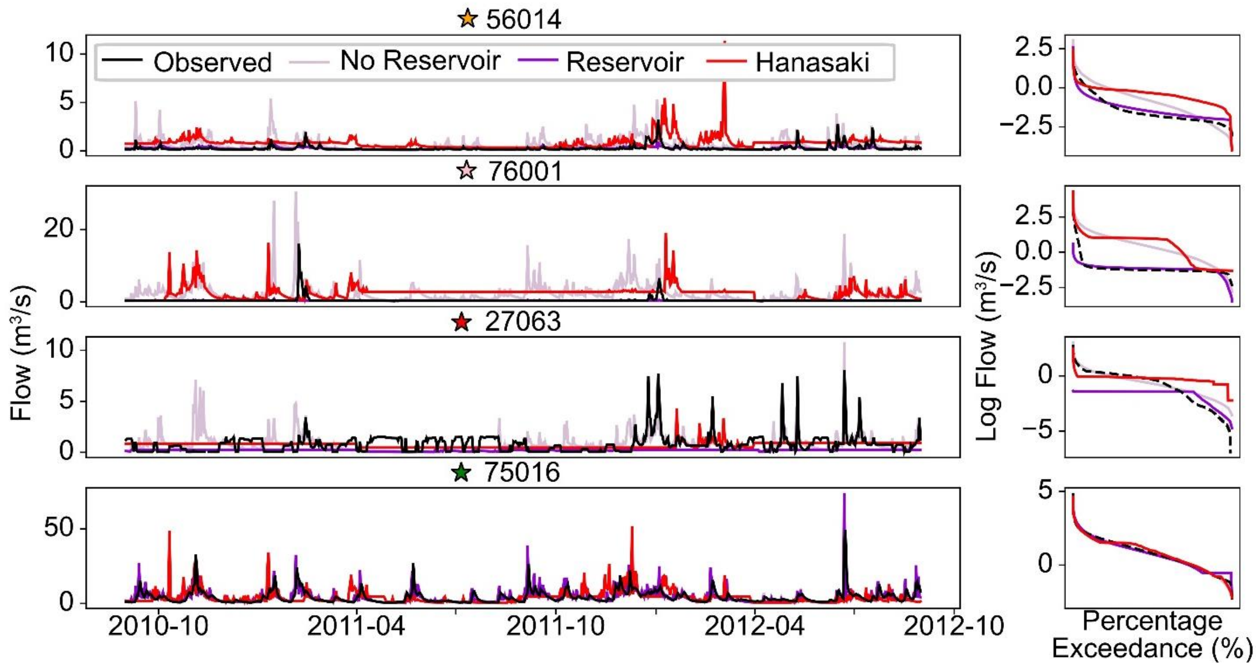

Figure S12. Hydrographs and flow duration curves top median simulation (nationally-consistent calibration) for selected reservoir catchments compared to simulations using the Hanasaki non-irrigation reservoir rules (Hanasaki et al. 2006).

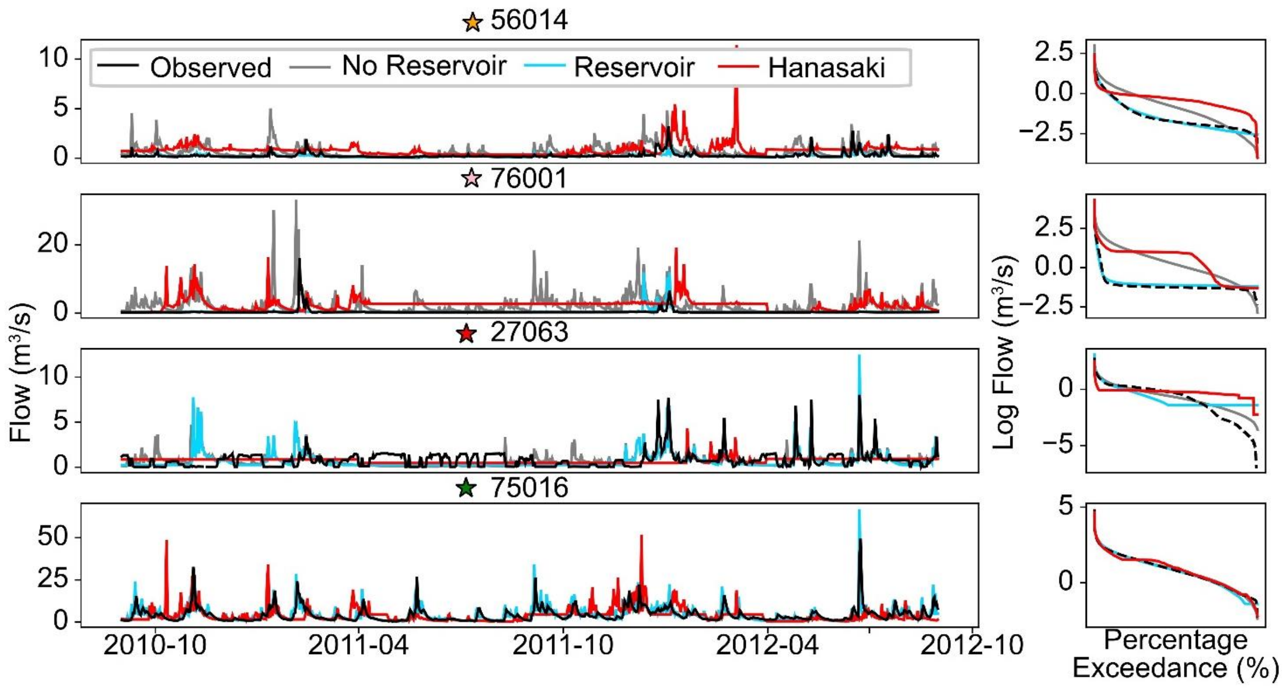

Figure S13. Hydrographs and flow duration curves from the top individual simulations (catchment-by-catchment) for selected reservoir catchments compared to simulations using the Hanasaki non-irrigation reservoir rules (Hanasaki et al. 2006).

Table S3. *Non-parametric KGE and associated components for simulations at four featured gauges using water supply operating rules (with both catchment-by-catchment (CBC) calibration and nationally-consistent (NC) calibration), no reservoir representation and Hanasaki et al. (2006) non-irrigation operating rules.*

Catchment	Operating rules	Water balance	FDC	Spearman's rank	Non-parametric KGE
56014	Water supply rules (CBC)	1.00	0.96	0.69	0.69
	Water supply rules (NC)	0.85	0.83	0.68	0.60
	No reservoir representation (NC)	1.93	0.89	0.67	0.01
	Hanasaki non-irrigation rules	1.93	0.69	0.40	-0.15
76001	Water supply rules (CBC)	1.04	0.92	0.45	0.44
	Water supply rules (NC)	0.63	0.62	0.43	0.23
	No reservoir representation (NC)	4.49	0.68	0.44	-2.54
	Hanasaki non-irrigation rules	4.49	0.63	0.35	-2.61
27063	Water supply rules (CBC)	0.83	0.80	-0.08	-0.11
	Water supply rules (NC)	0.24	0.70	-0.16	-0.42
	No reservoir representation (NC)	0.98	0.89	-0.17	-0.17
	Hanasaki non-irrigation rules	0.98	0.73	0.11	0.06
75016	Water supply rules (CBC)	1.01	0.97	0.89	0.88
	Water supply rules (NC)	1.01	0.94	0.87	0.86
	No reservoir representation (NC)	1.04	0.95	0.87	0.85
	Hanasaki non-irrigation rules	1.04	0.93	0.83	0.81

Minor Edits:

Line 53 and 55: I believe the abbreviation for GRanD should have a capital R as well.

Thank you for pointing this out, we have now corrected this across the manuscript.

Line 60: ResOpsUS contains over 600 dams. I would edit this line accordingly.

Thank you, we have corrected this.

Line 125 – 128: I would suggest adding parenthesis to equation 4 and 5 to denote the right side of the comparison.

Thank you, we have added this.

Figure 3: I would switch units in this figure from Ml/day to cubic meters per day since most audiences are more familiar with cubic meters.

We have changed these units.

Figure 6: The color difference between reservoir and observed is hard to distinguish. I would suggest altering the colors to be a bit more clear. I would also change the units of CF and ABS to be cubic meters /sec or per day so the units across all graphs are consistent. I do like that you kept the colors of the stars the same throughout Figure 4, 5 and 6.

We have altered the colors to make this clearer, thank you.

I like the colors on Figure 7 and the difference between the three curves is much easier to see. I would also change the units on CF and ABS to cubic meters per second or day for consistency.

We have changed these units.

Line 570: Could you give an example of a signature or test that you would suggest for other studies?

We have reworded this section to make it clearer that these new metrics still need to be developed. The text now reads:

We suggest that future studies should seek to develop new signatures which replace the correlation component of the KGE evaluation metric and can better capture behaviour in human-influenced catchments (Kiraz et al., 2023).

Line 589 – 590: Perhaps I missed a section, but I am not sure where the first half of this sentence comes from since the large scale hydrologic models that I am aware of all contain reservoirs in Great Britain. If the authors are referring to water supply reservoirs specifically, I would suggest rephrasing.

Here we are referring to national-scale hydrological models, we have altered this sentence to read:

The results of this study should encourage the inclusion of reservoirs in national-scale hydrological modelling across GB, since we have identified large gains in performance with minimal data and added complexity.

Although we acknowledge that there are several global-scale hydrological models which include reservoirs in GB, the resolution of such models is too coarse to inform water resources planning.