Dear editor, reviewers and Jan Seibert,

We would like to thank the reviewers and Jan Seibert for time they spent in providing such detailed reviews for our manuscript. A common theme amongst the comments was a need to consider more cases. To this end, please first see the general responses below.

General responses

GR1 – Case-study catchments

The availability and co-location of long-term hydroecological datasets represents the principle limiting factor in the application of the proposed modified covariance approach. This has been highlighted in both the methods and discussion sections of the revised manuscript.

A hydroecological dataset had recently been provided by the James Hutton Institute for a Scottish case study. Further, a scoping study of the English BIOSYS dataset and available hydroclimatological identified saw the identification of three further hydrologically diverse catchments. Additional factors we considered when identifying additional case studies were land use, BFI, location (north to south), seasonality and varied lengths of time-series. Thus, the consistency of the approach across a range of catchments is illustrated.

In addition to facilitating more general conclusions, this change allowed for the consideration of 40 distinct ER HIs covering all five facets of the flow regime. This represents a notable advantage over previous studies (Shrestha, Vis and Pool) which have been less diverse, focussing on a single suite of ER HIs and sub-basins of a larger catchment.

GR2 – Hydrological models

With respect to the number of hydrological models considered, the initial work on the River Nar did consider GR5J and GR6J. However, in stage 2, the models were invalidated. In this revision, this information has been included in the manuscript. Further, two out of the five catchments saw the validation of a small number of GR5J parameter sets.

Any further consideration of alternative model structures is deemed unnecessary (not commonplace in studies of a similar nature and beyond the scope of this work).

GR3 – Time-series length, parameter sets and equifinality

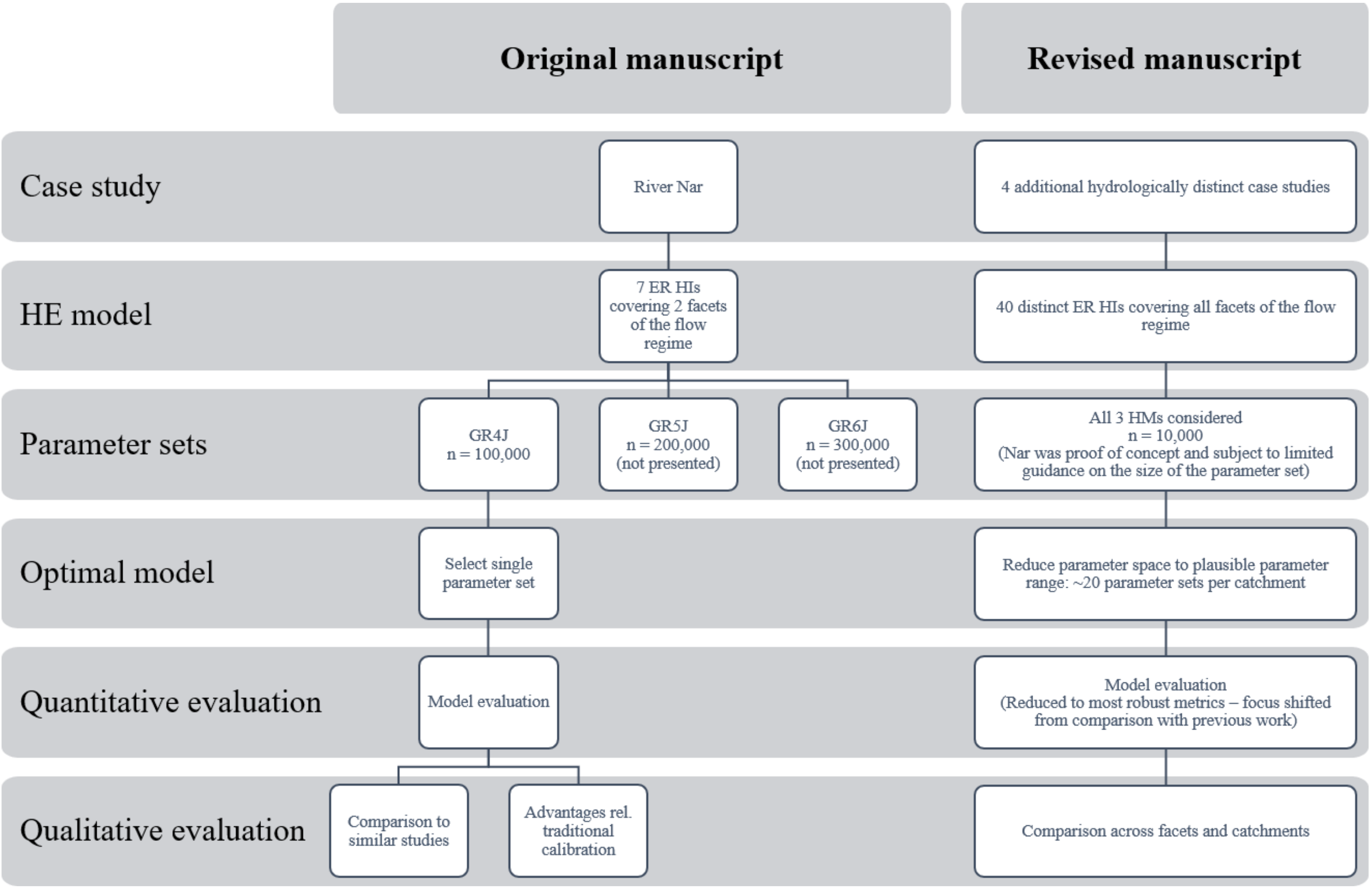
The validation of the model structure is achieved through comparison of the observed and simulated moments for covariance structure and ER HIs. This utilises the full length of the time-series. If observed and simulated moments are in agreement, the model is validated. To reduce the n parameter sets to a plausible parameter space, the log-threshold (limits of acceptability) is determined. For the parameter sets lying with this space, the time-series of annual ER HIs is subject to evaluation. In the original manuscript, a single parameter set was considered. In this revision, approximately 20 parameter sets per catchment (100 in total) are considered, serving to better highlight the performance and consistency across parameter sets.

The above revisions to the methodological application are also detailed in a figure at the end of this document. To reflect the changes in application, the introduction has been reworked to provide an overview of the development of methodological approaches. Additionally, the research questions have been refocused to a single aim, evaluating the performance of the modified covariance approach. We hope that these changes address your most pressing concerns and that you will reconsider this revised manuscript for publication.

Responses on a comment-by-comment basis are provided below (alternatively see attachment 2 for excel format), followed by the revised manuscript. Please note that the revised manuscript does not have tracked changes. These were removed due to the extent of the required changes (moving sections/figures, deletions and additions). If this is still required, we would be happy to provide a document with tracked changes.

Yours sincerely,

Annie Visser (Corresponding author)



Response to comments

Short comment – Jan Seibert

No.	(1) Comments from referees/public	(2) Author's response	(3) Author's changes in manuscript
1	Thanks for asking me for feedback for this interesting study on Researchgate. In general, I like this approach and think your work makes a good contribution towards better modelling of aspects in the hydrograph, which are relevant to hydroecology.	We thank you for your positive words towards the general concept of the presented approach.	-
2	My main concern with this study is that you used only one catchment. In our studies on the subject (Vis et al., 2015; Pool et al., 2017, as cited in your manuscript) we used 25 catchments and actually found that the performances differed among catchments. This means that there is a risk for somewhat random results if one uses only one catchment and more catchments would be advisable for robust results. At the very least this needs to be discussed, and it would be even better to extend your study to a few more catchments. Of course, handling all the simulations and their results can be painful as Marc (Vis), and Sandra (Pool) certainly will confirm.	Thank you for your suggestion. We agree and obtained the necessary hydroecological data to allow for the application of the methodology to four additional hydrologically diverse catchments.	See figure and general responses.
3	Isn't the covariance approach by Vogel and Sankrarasubramanian (2003, WRR) not also some form of calibration/fitting. Could you clarify the difference to traditional calibration a bit more?	The motivation for the consideration of covariances is to allow the hydrological model to be validated (or invalidated) prior to any attempt at parameterisation. Additionally, the approach allows for the identification of a plausible parameter space.	The closing paragraph of the introduction has been reworked to provide a better overview of the approach by Vogel and Sankarasubramanian. Further, the methodology has been more clearly delineated to improve clarity of the steps required. We hope these changes meet with your approval.
4	In the results, you report model parameters (P8 L7ff). Here it seems you derived one set of values. Later (P14) you discuss equifinality, but it was not fully clear to me how you considered equifinality in your study.	In the original manuscript we looked to the entire parameter space to validate the model. It was also indicated, though not applied, that additional parameter sets may be easily considered as all simulations/HEs are	The results from the a plausible parameter space which encompasses ~20 parameter sets per catchment are presented in this revised manuscript. An attempt to clarify the discussion of equifinality in the text has also

		considered in the application of the methodology.	been made (now in section 4.2.1). Additionally, see figure and general responses.
5	Also, I am not sure I understand your comment regarding the number of model runs (P14L15): why would your 100 000 runs scale to 400 000 runs for the HBV model? It seems like you are arguing with the number of parameters, 4 to 16 resulting in a factor 4. However, this should not be a factor, but the exponent resulting in a far larger increase of model runs to sample the parameter space with a similar density.	Thank you for this clarification. However, a subsequent sensitivity analysis using Sobol-sequencing highlighted that the number of parameter sets considered for the initial case study (River Nar, n = 100,000) was excessive. Subsequently, we were able to successfully reduce the number of parameter sets to a more manageable 10,000 without any loss of information.	This was no longer considered relevant, given the addition of four additional catchments, and has thus been removed in the revised manuscript for the sake of conciseness.
6	In the discussion (P11L14) you say that the replication of indices was good. Does this result apply for calibration or validation (both regarding time period but even more regarding index)? In our studies, we found in general the indices could be nicely (even perfectly) be replicated when calibrated, but performances decreased for the validation case.	We recognise that the data used in validation and subsequent parameterisation was not clear in the original manuscript.	This has been clarified in Section 2.4 (stages 1 and 2). We hope this change provides the necessary clarification.

Anonymous reviewer 1

No.	(1) Comments from referees/public	(2) Author's response	(3) Author's changes in manuscript
1	In this manuscript, Visser et al. evaluate the ability of the hydrological model GR4J to reproduce multiple hydrological indicators in a study catchment in south-eastern England. … etc. …	We thank you for your detailed and extensive review of the manuscript, it is very much appreciated. We hope that this revised manuscript addresses your concerns (specifically the 3 general comments).	See cover letter.

2	**General comments** The study aims at evaluating three research questions, which I think are very interesting. However, I have some concerns about the way the research questions are addressed: 1. Research question 1 is addressed by using one study catchment, one hydrological model, and one parameter set. To make more general conclusions I highly recommend to address the question by using many more catchments or multiple hydrological models. Given the current equifinality-paradigm in hydrology, I also recommend to account for uncertainty by using many parameter sets.	Thank you for your suggestions. We agree with the need for additional study catchments and obtained the necessary hydroecological data to allow for the application of the methodology to four additional hydrologically diverse catchments. We have now included three hydrological models across five catchments, n = 10,000 parameter sets for each, giving ~20 potential models per catchment.	We have made every effort to address your concerns through expanding the application and providing greater clarification in the text. See figure and general responses (cover letter) for overview.
3	2. Research question 2 about the comparison of various modelling studies is addressed by means of discussion. I am not sure if a discussion is enough to answer a research question. To me a research question should, if possible, be addressed by an analysis. If you wish to keep the comparison of your results with prior studies as a research question, I recommend to compare the studies in a quantitative way. Would it be possible that you contact the authors of the four studies to get access to more information?	-	In light of the feedback from reviewer 2, we have refocused the paper to consider a single aim. Thus, this research question has been removed; though some text remains in the discussion.

4	3. Research question 3 is again addressed by means of discussion without any explicit analysis. The goal of question 3 is to address the limitations of classical calibration such as i) effect of data uncertainty, ii) effect of thresholds applied to select behavioural parameter sets, and iii) effect of equifinality. I wonder if the current study set up allows to tackle these challenges. For example, it would be important that you could show that your proposed approach is less sensitive to disinformative data than other approaches. Or it would be helpful if you could show/ discuss in more detail how the selection of a threshold (limits of acceptability) in this study is different from other studies. And finally, it would be interesting to see how the proposed covariance approach reduces equifinality compared to other approaches.	Whilst this is no longer a research question, some text remains in the discussion. See column (3). With regards to disinformative data, we provide clear references to other work which has made these statements; therefore, it is our opinion that it is not necessary to consider this further. The use of limits of acceptability is inherently different to traditional approaches as there is no reliance on an optimisation algorithm and arbitrary objective function. Instead, a plausible parameter space is determined through explicit consideration of the statistical importance of each hydrological indicator. The aim of the approach is not to reduce equifinality, but rather to aid in the identification of a plausible parameter space. We recognise this was not clear in the original manuscript and hope that the revisions provide greater clarification.	In light of the feedback from reviewer 2, we have refocused the paper to consider a single aim. Thus, this research question has been removed; though some text remains in the discussion.
5	Specific comments 1. The authors motivate their research by stating that their approach is a step away from classical calibration. However, I don’t understand in which way the proposed approach is different from calibration. I agree that finding acceptable parameter sets by some kind of optimization algorithm is different from finding acceptable parameter sets by a Monte Carlo approach. However, the approaches only differ in the way parameter sets are generated, whereby both approaches require at some stage the selection of parameter sets by means of efficiency criteria. To me, both approaches can therefore be considered as model calibration. A noncalibrated model to me would be one where the 100’000 randomly generated parameter sets are used without making any further selection. To me it would be important that you come up with convincing arguments for the statement that the proposed covariance approach is not a (multi-objective) model calibration.	In the manuscript we state that the approach is different to traditional calibration which relies on optimisation algorithms and objective functions. The modified covariance approach does not use any optimisation algorithm, instead the entire parameter space is simulated and ER HIs derived. From this, the parameter sets are reduced to a plausible parameter range based on the statistical importance of the ER HIs. Additionally, validation is performed prior to parameterisation of the model.	We have rewritten the discussion of the approach in the final paragraph of the introduction and attempt to clarify section 2.4.1 in the methods. The reworking of the introduction more generally should also serve to place the approach in the wider context.

6	2. The topic of multi-objective calibration leads me to my next comment. It is multiple times mentioned (e.g. title or research question 1) that a covariance approach was used to determine the most suitable model parameters. If I understand correctly, the final selection of a parameter set is based on the combined evaluation of the covariance between precipitation and streamflow, the covariance between potential evaporation and streamflow, and seven hydrological indicators. I would therefore argue that it is not a pure covariance approach, but rather a multi-objective approach that includes covariance as one out of multiple efficiency criteria. Additionally, I think that covariance can be considered as a classical signature with the novelty that it is not a pure hydrological signature, but rather a hydroclimatic signature. To me, the very interesting part is the fact that the objectives (efficiency criteria) used in this multi-objective function are weighted by their importance. Concluding, I would recommend to replace the term “covariance approach” by a term such as “multi-signature approach” or “multi-objective approach”.	The method has been named to reflect Vogel and Sankarasubramanian (2003) where they referred to their method as the 'covariance approach'. Our modifications sees the consideration of hydroecological modelling outcomes through the limits of acceptability. Further, we feel that the method is at odds with the descriptions of multi-criteria approaches in other works. For instance, in their review of multi-criteria approaches, Efstratiadis and Koutsoyiannis (2010; https://doi.org/10.1080/02626660903526292) describe the use of algorithms and search techniques. To avoid confusion, and ensure consistency with the original method, we feel it is more appropriate to retain the title as is. Additionally, other studies focussing on replicating aspects of the time-series are not termed multi-criteria, e.g. Westerberg et al. (2011; https://doi.org/10.5194/hess-15-2205-2011). Additionally, it is worth highlighting that we have referred to this (in brief) as a modified covariance approach in the keystone paper of this project which is published in EMS: https://doi.org/10.1016/j.envsoft.2019.01.004	No change. Title has been updated to say "modified covariance approach" - this was an oversight on our part.
7	3. Is it correct that you select the final parameter set using the information of all 54 years? If yes, this would mean that you use the complete time series to find a parameter set and that there is no validation time period (meaning that all the error metrics are calculated for the calibration period). Since you have such a long time series, I would recommend to split the time series and use one part for an independent validation of the proposed approach.	We agree that split-sampling is the traditional approach - this was acknowledged in the original manuscript. However, one of the advantages of this approach is that validation and identification of a plausible parameter space is determined through covariances, thereby avoiding the need for splitting the time-series. Additionally, the other case studies do not have time-series as long as this (a necessary consequence of identifying catchments where sufficient hydroecological data is available).	Section 2.4 in the methods has been updated to clarify this point.

8	4. The “calibration” finally leads to the selection of a single parameter set. Why do you use only one parameter set? Is it because there was only one out of the 100’000 parameter sets that was behavioural? If there are multiple behavioural parameter sets I strongly recommend to use all of them. Otherwise, you will need very good arguments for putting all your confidence on a single model output.	The moments in the appendix (now figure A3) indicate excellent agreement between the observations and simulations. Indeed, if only a single parameter set was deemed suitable then it may call into question the validity of the hydrological model. In the original manuscript the validation did consider all 100,000. However, these were not all subsequently analysed in the evaluation.	In the updated manuscript we use the limits of acceptability to identify a plausible parameter space of ~20 parameter sets that minimises the covariance error (this is not fixed - it is simply for illustrative purposes). Consideration of these 20 parameter sets further allows for expanded discussion of consistency.
9	5. As far as I understood, hydrological indicators were calculated for each single year. Given that hydrological indicators were shown to be only robust if calculated over many years, how do you think this influences your results? Do you think that the yearly variability of the indicators can obscure/influence the uncertainty coming from the approach?	Validation and identification of the feasible parameter space looks at the full length of the time-series. Model evaluation looks at annual indicators as their projection represents the overall modelling objective.	An additional comment has been added to the close of the paragraph in section 4.2 Advantages and limitations of the modified covariance approach.
10	6. I think it would be worth to spend some time in adapting the introduction. For example, the two first paragraphs are very generic and I am not sure how much information they contain related to your research questions. You could shorten these paragraphs to one/two sentences and then extend the introduction to provide more background on e.g. multi-objective calibration or other studies modelling hydrological indicators.	-	We appreciate this suggestion and have adapted accordingly. The introduction is now more focussed on background and the development of methods over time.

11	Detailed comments 1. Abstract: You mention in the abstract that one benefit of the proposed approach is the reduction in overall time-demands. Could you specify what exactly you are thinking of? The first thing that comes into my mind is that you want to reduce computational time. However, GR4J is a model, which is not very demanding in terms of computational effort. I was therefore wondering if you thought about using a more time-consuming model to proof that the approach does save computational time. This would also need a comparison of a traditional approach to your approach, which, I know, will need quite some time. Of course, you could also just explain or weaken your statement.	This was indicated in the original manuscript (in revised manuscript it is the last paragraph of section 4.2.1 General advantages). However, this does represent an interesting opportunity for future work - a comparison of hydrological models in terms of computational time, performance and consistency.	Removed from abstract. Only consideration now is in the discussion.
12	2. Study area: You mention that the catchment is an SSSI, that there is significant pressure on the river, and that the river has a highly seasonal flow regime. I think it would be interesting to add a sentence or two saying why the catchment is an SSSI, what kind of pressure sources exist, and how the seasonality looks like (mostly winter streamflow?).	The seasonality is captured in the ER HIs. With the addition of four catchments, the level of detail on each catchment has been reduced to a level reflecting similar studies (summary table + map).	-
13	3. Fig. 1: The markers for River Nar and Lexham village are difficult to differentiate in the figure.	This level of detail was no longer necessary, figure is thus redundant and has been removed.	-
14	4. Fig. 2: I was wondering why you decided to show a Figure of the model structure of GR4J? Given that you don’t compare multiple models with different structures or that you don’t extensively discuss model parameter values, I think you could remove the figure.	-	This has been moved to the appendix. Description of the additional hydrological models has also been added.
15	5. Table 2: I suggest to name the last item of the header “relative importance”.	This was not relative importance as it had not been standardised to a range of zero to one. However, the Table has been removed as it was no longer relevant.	However, it was observed that in the text the term relative had been used incorrectly towards the beginning. This has been corrected for clarification.
16	6. P6 L18: If I am not wrong, the reference to Fig. A2 comes before the reference to Fig. A1. So maybe you could switch the position of these two figures in the appendix.	-	This has been addressed in the revised manuscript.

17	7. P7 L4: Could you say specifically which error you minimise between observed and simulated covariance and HI? Is it the percent error?	-	Text updated: "the ability to replicate or minimise the error (percentage difference)"
18	8. Fig. 4: 1) The y-axis label is "percent error" while the legend says "difference between observed and simulated values" – what is correct? 2) The plot location of the hydrological indicators on the x-axis does not fully agree with the values in Table 2, e.g. Q70Q50 has according to Table 2 a low relative importance while it has a high one in the figure. 3) You use this figure two illustrate the concept of the limits of acceptability and to show the result of the best parameter set. Is there a way you could separate methods and results part in this figure?	1) No longer relevant and clarified in #17. 2) No longer relevant. Error was the result of mislabelling of the points in the figure. 3) See column (3).	The figure has been adjusted to be a simple exemplar which does not use real data.
19	9. P8 L6: The first reference you do in the results part is to a figure in the appendix (Fig. A2). Given the importance of this figure, I would suggest to have it in the main body of the manuscript.	-	This figure is now only considered as an exemplar of how the moments may be visualised (methodology).
20	10. P9 L7: You mention that you evaluate the model in terms of performance and consistency. I would therefore recommend that you rearrange the results chapter do this evaluation in a very clear way.	The terms performance and consistency are introduced in order to provide clear definition of terms. In the original manuscript, consistency was with respect to performance across the seven ER HIs. We appreciate your suggestion in that context.	In this revised manuscript, we look for consistency across parameter sets and catchments as well. With a view to minimising the length of the results section, results are presented in sub-sections relating to each of the tests applied. The tests are considered in the order they are applied, thus the addition of each test serves to advance the 'story'.
21	11. P9 L25-28: I would delete the first two sentences of this paragraph because they are methods and not results. I would also delete the last sentence of this paragraph and add the reference to Fig. 8 somewhere in brackets.	Thank you for highlighting this. The latter is no longer relevant in the revised manuscript.	-

22	12. Figure captions: Could you be more specific in the figure captions, i.e. could you for each figure say how many data points are in each plot? I think it is important to guide the reader by telling if e.g. a histogram contains 54 simulation years or n parameter sets.	-	In the revised manuscript 100 parameter sets are considered for 5 catchments (~20 per catchment). Section 2.5 model evaluation has been updated to clarify this point with an example for the River Nar included. This statement is true for all metrics applied. Results are presented as box plots to illustrate the range of values across the parameter sets.
23	13. Fig. 5 and Fig. 6: These threes (sub)plots contain very similar information. I recommend to find a way to condense the information into a single figure. In Fig. 5a, what do the numbers in the brackets of the header mean?	-	In this revised manuscript the results from the statistical tests are presented in a tabular format. Given the size of the required table (with the increased number of catchments and ER HIs) this table is located in the appendix.
24	14. Fig. 6: The figures contain a relatively small number of points. I was wondering if you can merge the three figures or if a table/ heatmap would be more suitable to show the results?	-	As above in #23.
25	15. Fig. 8: The figure does not contain a dot for the 0-25 quantile of RevPos. I would suggest that you mention the reason for that in the figure caption and not somewhere in the main text. Maybe you could also plot the dot at the margin of the figure together with an arrow indicating that it is an outlier.	We thank you for this comment and will certainly bare this in mind for future work. In this instance, the consideration of additional catchments have seen this figure evolve considerably.	Not relevant. Figure replaced.
26	16. P21 L11: The reference to Cramer is not at the correct location and is lacking a year.	-	To ensure consistency, all references which were located within Table 3 have been moved to the main body text.

Anonymous reviewer 2

No.	(1) Comments from referees/public	(2) Author's response	(3) Author's changes in manuscript
1	This manuscript presents a modified approach after Vogel and Sankarasubramanian's covariance approach and extend the scope from a single-variable problem to a multiple-variable problem. The authors found that the approach can reduce model uncertainty and also time consumption. Overall, I think this is a great idea and the proposed approach and the results are a valuable contribution to the hydrological community. I recommend its publication after the following comments are addressed.	We thank you for your feedback and suggestions. We hope the revisions meet with your approval.	Detailed below and see also cover letter.
2	General comments: 1. The work is very site-specific and model-specific, which needs to be tackled or at least acknowledged.	-	An additional four hydrologically diverse catchments have been considered (this is necessarily limited due to hydroecological data availability). We have expanded the paper to explicitly consider GR4J, 5J and 6J (these were considered in the original application but were invalidated). Any further consideration of alternative hydrological model structures is not deemed necessary as similar studies have not accounted for this.
3	2. Research Question 1 is addressed in the manuscript with direct, quantitative analysis but Questions 2 3 are not. My recommendation to the authors is to (1) provide more in-depth analysis for these two questions or (2) modify their research questions in Introduction. I feel that Question 1 can be listed as the single research question of this work, whereas your questions 2 and 3 can be raised in the Discussion section.	-	We agree; additionally, the consideration of additional catchments and parameter sets allows for more generalised conclusions and therefore the second and third research questions are no longer necessary. Where relevant, the text in the discussion has been retained.

4	3. I think it helps the readers a lot if the authors can provide some clarification on how the 54 years of data were used in their approach. Was there any split? If so, which part is used as calibration and which part for validation? Why were HIs calculated for each year?	Validation and identification of the feasible parameter space looks at the full length of the time-series. Model evaluation looks at annual indicators as their projection represents the overall modelling objective.	Clarification has been provided in section 2.4 Covariance approach with regards to the time-series. An additional comment re annual indicator uncertainty has been added to the close of the paragraph in section 4.2 Advantages and limitations of the modified covariance approach.
5	4. There can be some relocation of the figures and tables. For example, I think Figure 2 and Table 1 can be moved to Appendix or Supporting Information, since the hydrological model is already published and it is not the goal of this work to investigate the model itself. To the contrary, some of the Appendix information is critical and should be placed in the main text, e.g., Figure A1, A2, Table B1.	Figures and tables have been relocated. Figure A1: No longer considered relevant due to the shift in focus of the paper. To account for this behaviour we adapted the HAF code for integer values. Figure A2: This remains in the appendix as it now represents only an exemplar for one of the case studies (it is not practically possible to display these figures for all catchments and hydrological models, 5*3). Table B1: With the adjustment of the research questions this table has been removed, the level of detail is no longer required.	-
6	Specific comments: 5. P1L9: Does Vogel and Sankarasubramanian's covariance approach need a citation in the abstract? At least the journal and year of that publication should be provided.	-	This modification has been made; it as the discretion of the editor/journal whether they will allow it (in my experience the removal of a citation has been requested).
7	6. P1 Introduction: The first two paragraphs can be largely shortened. Some of the details are well known and may not be necessary.	-	The introduction has, generally, been revised to reflect the development of methods over time.
8	7. P2L30: Can you be more specific on "many of these problems"?	-	This line has been deleted as the research questions have been modified.

9	8. P3L29: By definition, Q90 should be larger than Q10. So Q95 below should be Q5.	We are aware that Q90 can mean low and high flows interchangeably. In the context of ER HIs, Q90 is the flow exceeded 90% of the time, hence, a low flow. See Shrestha et al., 2014, Vis et al., 2015 and Pool et al., 2017 for example.	-
10	9. P5: The “IT approach” might be presented with more details.	-	Section 2.3 Data has been expanded to consider the hydroecological model development further. Additionally, further details on information theory are provided in the Appendix.
11	10. P6: Figure 3 may be modified by highlighting the boxes or flows that represent your new approach, as compared with the original 2003 approach.	We agree this would be useful, however the level of difference is marginal. In stage 1 and 2 the only difference is that more than one HI is determined. The principle difference comes in the third stage and the identification of the plausible parameter space.	A new line has been added to the end of paragraph one in section 2.4 Covariance approach.
12	11. P6L16: Check for consistency in the tense of verbs.	-	Addressed as appropriate.
13	12. P6L18: Fix "the of the"	-	Addressed.
14	13. P7: Figure 4, A dashed horizonal line at 0.0 may be added.	-	The revised figure based on an exemplar now depicts error.
15	14. P8: Table 3, it is not readily clear to me what “p>0.05” represents.	-	Added clarification in text and caption that this refers to the significance 'threshold' specified. It has also been updated to 0.001.
16	15. P9: Some of the context in the Results section should belong to Methods, e.g., the description of HAF	-	Addressed as appropriate.
17	16. P9L21: “précised”?	Summarised.	Not present in revised text.
18	17. P11: Note the error “increas” in Figure 8 caption.	-	Addressed.
19	18. P11: The first three subsections of Section may contain a bracket noting the correspondence to the Research Questions.	-	Not relevant in the revised text.
20	19. P19: Note the error in Table B2 caption.		Not relevant in the revised text.

Replication of ecologically relevant hydrological indicators following a modified covariance approach to hydrological model parameterisation

Annie Visser[1], Lindsay Beevers[1], Sandhya Patidar[1]

[1]Institute for Infrastructure and Environment, Heriot-Watt University, Edinburgh, EH14 4AS, UK

Correspondence to: Annie Gallagher Visser (av96@hw.ac.uk)

Abstract. Hydrological models can be used to assess the impact of hydrologic alteration on the river ecosystem. However, there are considerable limitations and uncertainties associated with the replication of the required, ecologically relevant hydrological indicators. Vogel and Sankarasubramanian's 2003 (Water Resources Research) covariance approach to model parameterisation represents a shift away from traditional algorithmic calibration-validation focussed on objective functions. Using the covariance structures of the observed input and simulated output time-series, the plausible parameter space, the region of parameter space which best captures (replicates) the characteristics of a hydrological indicator, may be identified. In this study, a modified covariance approach is applied to five hydrologically diverse case study catchments with a view to replicating a suite of ecologically relevant hydrological indicators identified through catchment-specific hydroecological models. The identification of the plausible parameter space (here n ≈ 20) is based on the statistical importance of these indicators. Evaluation is with respect to performance and consistency across each catchment, parameter set, and the 40 ecologically relevant hydrological indicators considered. Timing and rate of change indicators are the best and worst replicated respectively. Relative to previous studies, an overall improvement in consistency is observed. This study represents an important advancement towards the robust application of hydrological models for ecological flow studies.

1 Introduction

Increases in societal water demand and climatic variability raise questions over the long-term sustainability of water resources (Gleick, 1998; Klaar et al., 2014; Davis et al., 2015; Gleick, 2016). As the ecological role of flow is better understood, it has become widely acknowledged as the major determinant of the ecological health of the riverine ecosystem (e.g. Power et al. (1995); Lytle and Poff (2004); Arthington et al. (2006)). Consequently, changes to flow threatens both the ecological health of rivers and their ability to provide the vital ecosystem services upon which humans depend (Vörösmarty et al., 2010; Arthington, 2012).

Beginning in the late 1940s in the United States, the need to balance the conflicting demands of both human society and those of the ecosystem saw the emergence of the environmental flow movement. Environmental flows have been defined under the Brisbane Declaration (2007) as: *"…the quantity, timing, and quality of water flows required to sustain freshwater and* *estuarine ecosystems and the human livelihood and well-being that depend on…"*. Tharme (2003) documented that over 200 formal environmental flow assessment methods had been developed.

Quantifying the relationship between flow and ecology is pivotal for the determination of environmental flows (Bunn and Arthington, 2002; Arthington et al., 2006; Poff et al., 2010; McManamay et al., 2013). Richter et al. (1996) identified five facets of the flow regime required to support the riverine ecosystem: magnitude, frequency, duration, timing and rate of change. Alteration of the flow regime invariably leads to significant ecologic change. To date, over 200 ecologically relevant hydrologic indices (ER HIs) have been proposed (Olden and Poff, 2003; Monk et al., 2006; Thompson et al., 2013). Poff et al. (2010) and Peters et al. (2012) each describe environmental flow frameworks, which call for the determination of ER HIs via hydrological model simulations of flow. At the time of publication (of these frameworks), the application of hydrological models for the determination of ER HIs was in its infancy (Knight et al., 2011). Indeed, early work was largely based on regional statistical approaches which had been in use since the 1960s in the United States (for the determination of water resource relevant HIs; for example see Knight et al. (2011) and Carlisle et al. (2010)). Murphy et al. (2012) compared such ER HIs against those determined from simulated flows, finding that, without targeted calibration to specific HIs, *"the widespread application of general hydrologic models to ecological flow studies is problematic"* (p. 667). However, such statistical approaches are unsuitable when assessing the impact of hydrological change on the river ecosystem (e.g. as a result of engineering intervention or under a changed climate) or for the simulation of ecological flows in ungauged catchments. A hydrological modelling approach is thus necessary.

Model performance and consistency are watchwords for this study. After Euser et al. (2013), model performance is defined as the ability to mimic the behaviour of catchment hydrological processes; consistency represents the ability of the hydrological model to reproduce a suite of ER HIs across parameter sets, hydrological models and catchments.

Significant bias has been observed in hydrological models calibrated following traditional approaches (based on the use of objective functions; Grayson and Blöschl (2001); Blöschl and Montanari (2010); Westerberg et al. (2011); Pushpalatha et al. (2012)). In context, when evaluating the suitability of model simulated HIs (six water resource relevant HIs and 32 ER HIs), Shrestha et al. (2014) observed that the water resource relevant HIs were well-replicated whilst notable differences were observed for the ER HIs related to the facets of the flow regime duration and rate of change. Informed by recent advances in hydrological modelling more generally (Seibert, 2000; Efstratiadis and Koutsoyiannis, 2010), Vis et al. (2015) compared the ability of single and multi-criteria objective functions to replicate twelve ER HIs. The best performance was achieved with multi-criteria objective functions, though a consistent negative bias was observed. Despite these advances, overall performance was inconsistent, being dependent upon the ER HI considered. Blöschl and Montanari (2010) observed that the reliability of hydrological modelling approaches which try to 'model everything' is analogous to simply 'throwing the dice'. To address this, they call for a move towards simpler models, tuned to focus on specific characteristics of the flow regime; successful applications of such an approach include Westerberg et al. (2011). Most recently, Pool et al. (2017) considered an array of multi-criteria objective functions using Nash Sutcliffe Efficiency (NSE) and 13 ER HIs. Overall, results were positive, with

ER HIs generally well-replicated in calibration; validation performance was subject to greater variability. Those ER HIs not explicitly included in the objective function exhibited greatest uncertainty.

The past ten years has seen the replication of ER HIs evolve from statistical approaches to single and multi-objective rainfall-runoff modelling. Whilst improvements have been notable, to date no approach has been able to achieve performance and consistency concurrently, raising questions as to whether these approaches are able to achieve the 'right answer for the right reasons'. Further, Pool et al. (2017) highlight two points which remain unaddressed: (1) a need to determine which ER HIs are relevant in order to guide model parameterisation; and (2) laborious recalibration of the hydrological model is necessary if the suite of HIs is changed. In this paper we look to redress these limiting factors through the application of a modified covariance approach. The objective of Vogel and Sankarasubramanian's (2003) covariance approach is to identify the plausible parameter space which captures (replicates) the characteristics of a specified HI. This is achieved by focussing on the ability of the hydrological model to capture the observed covariance structure of the input and output time-series. The use of covariance relationships in this way is not new, with examples including the modelling of ice sheets (Wu et al., 2010) and ocean salinity (Haines et al., 2006). Vogel and Sankarasubramanian's covariance approach is limited by its focus on a single HI, preventing its use for the determination of a suite of ER HIs. This paper builds on the covariance approach, adapting the methodology to consider a suite of ecologically relevant hydrological indicators; the determination of these ER HIs is based on the outcomes of hydroecological modelling using an Information Theory approach. To determine the ability of the modified covariance approach in replicating ER HIs, the method is applied to five case study catchments across the UK using the daily models from the GR (Génie Rural) suite of hydrological models (GR4J, GR5J and GR6J, 4-6 free parameters; Coron et al. (2018)).

2 Methods

2.1 Study areas

To illustrate the generality of the modified covariance approach, it is necessary to apply the proposed methodological approach to a range of catchments with differing characteristics (Andreassian et al., 2006; Gupta et al., 2014). The UK is home to a wide range of hydrological environments, with 18 different river types specified under the WFD (Rivers Task Team, 2004). Under the modified covariance approach, hydroecological models inform the parameterisation of the hydrological model. In the past, a mismatch between the co-location of sampling sites as well as the length of time-series represented limiting factor (Monk et al., 2006; Knight et al., 2008). In the UK, this is avoided by the recent publication of the UK BIOSYS archive (long-term ecological monitoring data from across the UK; Environment Agency (2018)). In this study, we consider a total of five catchments across the UK, from the north of Scotland to the south-west of England (Fig. A1), of varying size, altitude, baseflow index (BFI) and land use; catchment characteristics are summarised in Table 1.

Table 1. Summary of case study catchment characteristics.

		Tarland Burn	River Trent	River Ribble	River Nar	River Thrushel
Flow gauge and catchment	*Location*	Aboyne	Stoke-On-Trent	Arnford	Marham	Hayne Bridge
	Longitude	-2.7758	-2.1624	-2.2471	0.5472	-4.2424
	Latitude	57.0777	53.0175	53.9962	52.6783	50.6584
	BFI	0.66	0.44	0.25	0.91	0.39
	Drainage area (km^2)	70.9	53.2	204	153	57.6
	*Principal land use**	Mountain, heath and bog	Urban and grassland	Grassland	Arable and horticulture	Grassland
Data	*Years*	2003-2016	1989-2016	2000-2016	1961-2015	1989-2016
	Flow data source	JHI (2018)	NRFA (2018)			
	Climate data source		Met Office (2018a) and Met Office (2018b)			

2.2 Hydrological model

The principle of parsimony, known as Occam's razor, posits that a solution should be no more complex than necessary. In the context of hydrological modelling, model simplicity relative to performance is thus made key (Kokkonen and Jakeman, 2002; Perrin et al., 2003; Beven, 2012). To this end, the three lumped models from the GR-J series of daily hydrological models was selected (Perrin et al., 2003): GR4J, GR5J and GR6J (4, 5 and 6 free parameters respectively; Perrin et al. (2003); Le Moine (2008); Pushpalatha et al. (2011)). The GR-J series of models have been applied in a variety of hydrological contexts, examples include: Perrin et al. (2008); Coron et al. (2012); Smith et al. (2012); Coron et al. (2017).

The three models are based on soil moisture accounting (Fig. A2). The model inputs, *P*, the catchment rainfall depth, and *E*, the average depth of (potential) evapotranspiration), fill the production store with a capacity of *x1* mm. The routed depth of water, *Pr*, is determined by the rate of percolation, *F(S, x1)*, as well as water in excess of the storage capacity. To simulate the time difference between rainfall event and flow peak, *Pr* is divided into two flow components and routed through unit hydrographs, time base *F(x4)* days. Finally, the groundwater exchange term *gw*, *F(x2)*, acts on the routed, *Qr*, and direct flow, *Qd*, components; a positive value indicates inflow from groundwater whilst a negative represents water export. The total flow, *Q*, is determined by summing the routed and direct flows. With the aim of improving general modelling efficiency (Anderson Michael et al., 2004; Hughes, 2004), GR5J sees the addition of functions representing the interaction between channel and aquifer flows (Le Moine, 2008); the corresponding free parameter, *x5*, represents the inter-catchment exchange threshold which acts at the same points as *x2* (groundwater exchange). With a view to improving low-flow simulations specifically (Pushpalatha et al., 2011), the GR6J model sees the addition of a parallel store with capacity *x6*. The models are applied using the R package *airGR* (Version 1.0.15.2; Coron et al. (2017); Coron et al. (2018). Parameter limits are summarised in Table A1.

2.3 Data

The ER HIs were determined based on the outcomes of hydroecological modelling for each catchment. Following Visser et al. (2018a), the hydroecological model was developed using multiple linear regression with an information theory approach; a summary of the modelling approach is provided in Appendix A.2. The information theory approach provides a measure of the statistical importance of each ER HI (measure of the statistical weight of evidence for the inclusion of the index in the model) in addition to minimising and quantifying uncertainties (structural and parameter). Consequently, more conclusive statements may be made with regards to the model and the relevance of the selected ER HIs. Definitions of the ER HIs included in the hydroecological models, and their importance, are available in Table B1. To reflect seasonality in the flow regime, the indices are differentiated by hydrological season: winter (ONDJFM) and summer (AMJJAS). A summary of the distribution of the ER HIs per facet of the flow regime, season and river is provided in Table 2.

Table 2. Number of ER HIs per facet of the flow regime, season (W and S denote summer and winter respectively) and river. Sum totals are detailed in the final columns and rows.

		Tarland Burn		River Ribble		River Trent		River Nar		River Thrushel		Sum per facet
		W	S	W	S	W	S	W	S	W	S	
Magnitude	*Statistic*	1	1	1	2	1			1	2		9
	Ratios – Log quantile						2	1			1	4
	Ratios – Median-quantile				4	2			3	1	2	12
	Monthly	2				1				1		4
Duration			2	1		2				1		6
Frequency		1		1	1	1	1			2		7
Timing			1		2	2				1		6
Rate of change					1		1	1	1	1		5
Sum per season per river		4	4	3	10	9	4	2	5	9	3	53

Continuous (daily) time-series of mean flow, precipitation and potential evapotranspiration serve as model input; flow and climate data availability are summarised in Table 1 previously. Potential evapotranspiration was estimated using a temperature-based PE model (Oudin et al., 2005).

2.4 Covariance approach

The covariance approach was developed by Vogel and Sankarasubramanian (2003), where the aim was to replicate a specific HI rather than the flow time-series. The modification of the covariance approach in this study allows for the consideration of a suite of ecologically relevant HIs. The modified covariance approach is implemented over three stages (Fig. 1); stages 1 and 2 are as in Vogel and Sankarasubramanian (2003), with the exception that multiple ER HIs are calculated, with the final stage representing the modification introduced in this study.

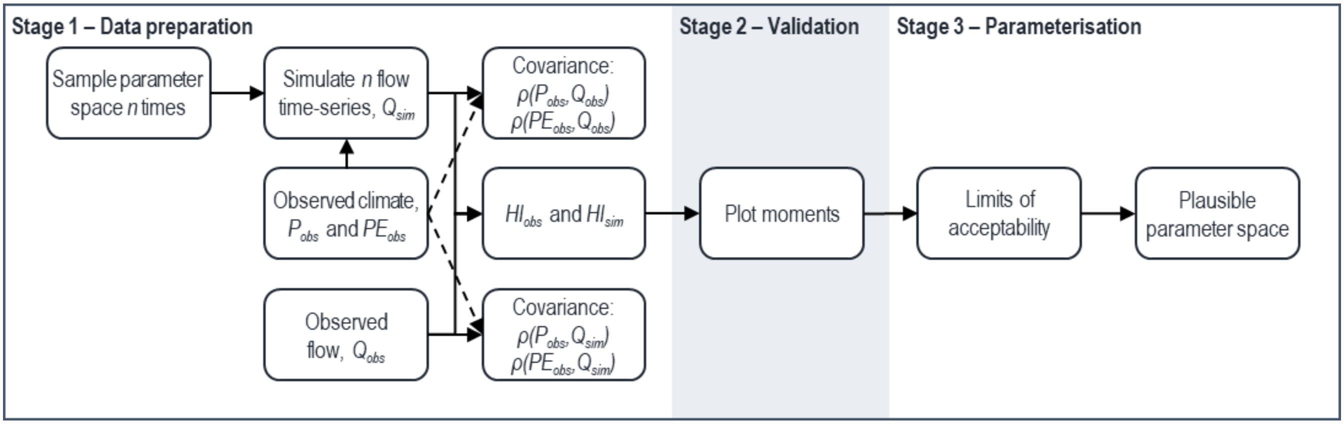


Figure 1. Overview of the three stages of the modified covariance approach to model parameterisation.

Stage 1, data preparation: The parameter space of the three hydrological model structures was sampled within the limits specified in Table A1. With a view to addressing both parameter sensitivity (Tong and Graziani, 2008; Wu et al., 2017) and the number of parameter sets considered, the parameter space was sampled uniformly based on Sobol quasi-random sequences (a Quasi-Monte Carlo method). The River Nar catchment served as the 'proof-of-concept', consequently, for this catchment, 100,000, 150,000 and 200,000 independent parameter sets were selected for the GR4J, GR5J and GR6J hydrological models respectively; for the remaining four catchments, 10,000 parameter sets were considered (per hydrological model).

For each parameter set, flow time-series were simulated based on the full time-series of the observed climate data. For each of these flow time-series, a corresponding set of covariances (between observed climate and simulated flow) and HIs were computed. The observed covariance and HIs are also determined.

Stage 2, validation: Under the traditional approach, the hydrological model is validated following calibration using an optimisation algorithm; this presupposes that the selected hydrological model is suitable. However, with a covariance approach, the model structure is validated prior to parameterisation. The model is validated when the observed moments lie within the simulated moments (sampled parameter space). This may be facilitated through plots of the observed and simulated relationship between the (a) covariance between precipitation and flow, $\rho(P,Q)$, and HIs; and (b) covariance between potential evapotranspiration and flow, $\rho(PE,Q)$, and HIs. If the moments do not agree, the model is invalidated. An exemplar for the River Nar is provided in Fig. A3. The moments may also be used to determine model equifinality (the existence of multiple behavioural parameter sets; Beven (2006); Efstratiadis and Koutsoyiannis (2010)).

In addition to the above, validation of the hydrological model in this way avoids the need for split-sampling (calibration and validation time-periods), thereby allowing the use of the full length of the hydroclimatological time-series in both the validation and subsequent parameterisation.

Stage 3, parameterisation: Selection of a model parameter set was based on a specified limit of acceptability (summarised in Fig. 2), i.e. the ability to replicate or minimise the error (percentage difference) between the observed & simulated covariance structures and ER HIs. In Vogel and Sankarasubramanian (2003) the focus was on the replication of a single index, whilst, in this study, the objective was the replication of multiple indices. To this end, a limit of acceptability was specified per index, with each ER HI assigned maximum error threshold based on their normalised or relative importance. The ER HI importance (Table B1) was normalised (rescaled to a range from zero to one) per catchment and the covariances assigned a relative importance of one, equal to the most important index. The catchment specific limits of acceptability were specified as the relationship between the relative importance and a user-specified allowable error range. If no parameter sets are selected, the model structure is invalidated.

In this study, an exponential model of the form $y = e^{mx+c}$ was specified for each catchment. In order to account for equifinality, the maximum error was set such that the feasible parameter space was limited to approximately 20 distinct parameter sets. In Fig. 2 an exemplar is presented where the limits of acceptability are adjusted with a view to identifying the plausible parameter space where n = 3.

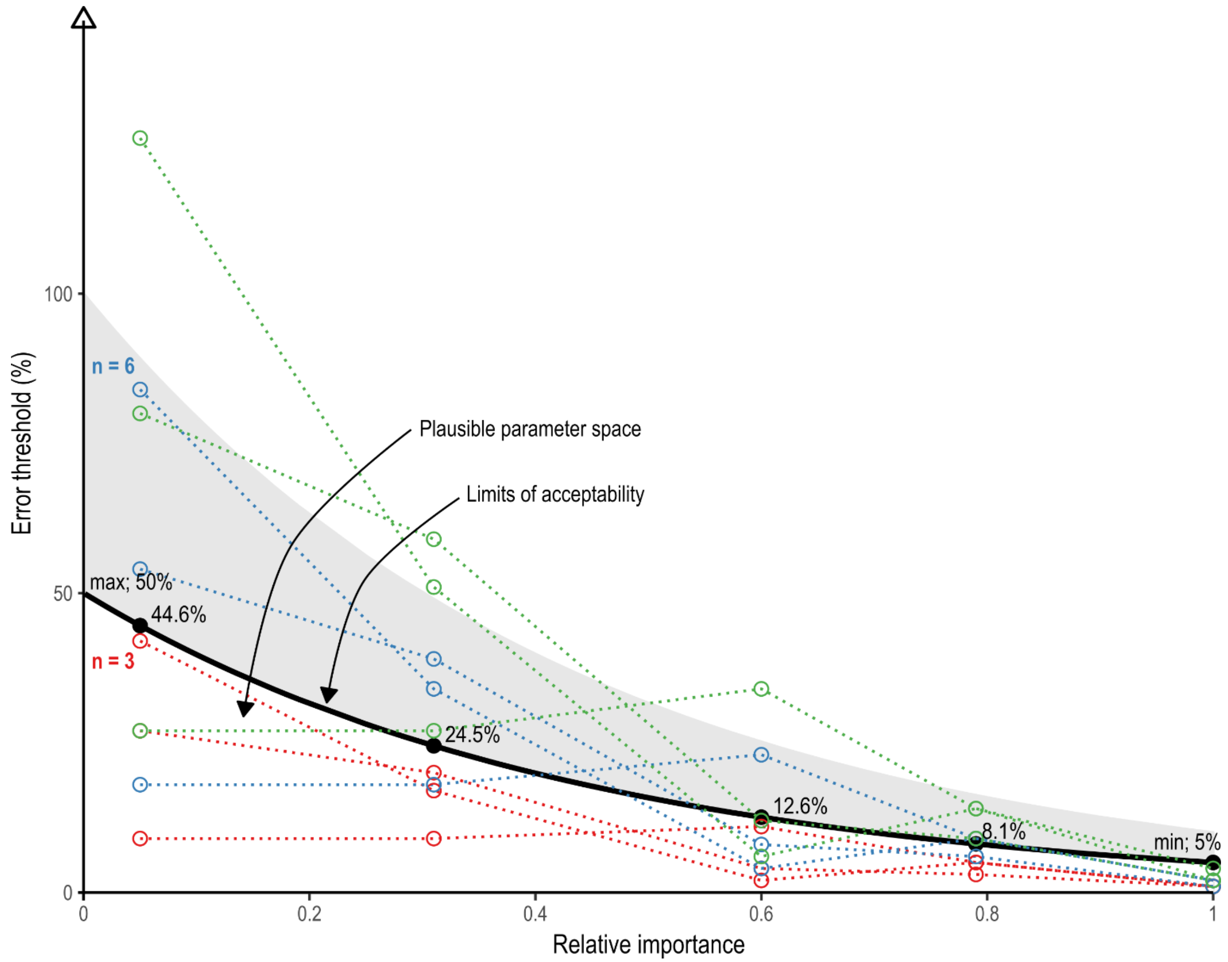


Figure 2. Conceptualisation of the limits of acceptability, depicted here as the log-linear relationship between relative importance and the allowable (absolute) error thresholds per indicator and covariance. The limits of acceptability are reduced until *n = 3* parameter sets lie within the plausible parameter space. In this example, the error threshold ranges from 5%, where the relative importance is one, to a maximum of 50%. The maximum allowable error per example indicator is marked.

2.5 Model evaluation

The ability of the parameterised models in replicating the ecologically relevant hydrological indicators was evaluated through the evaluation metrics detailed in Table 3 (determined with reference to prior studies with similar modelling objectives: Shrestha et al. (2014); Vis et al. (2015); Pool et al. (2017)). Metrics were determined across the full time-series for each catchment~parameter set (e.g. for the River Nar, 54 years of seasonal ER HIs were determined for each of the 23 parameter sets). Three statistical tests were applied, where the goal is the rejection of the null hypothesis (α = 0.001). Welch's t-test considers the correlation between the means of the observed and simulated indicators, whilst the KS and CvM (Cramér, 1928; Anderson, 1962) tests look to the distribution of the interquartile range and tails respectively; agreement indicates a relationship between the observed and simulated ER HIs. The hydrologic alteration factor (HAF) is adapted from the IHA approach (Mathews and Richter, 2007). It is a measure of the simulated and observed frequencies of values within three target percentile ranges: 0-25th, 25-75th, and 75-100th. As a measure of distribution, HAF is essentially a simplification of the distribution function. The acceptable range of HAF values is defined as ±0.33. Finally, two measures of error are determined: model efficiency, or the Nash-Sutcliffe Efficiency criterion (NSE), and the mean arctangent absolute percentage error (MAAPE), designed to address the limitations inherent to mean absolute relative error (Kim and Kim, 2016).

Table 3. Descriptions, definitions and optimal values for the applied evaluation metrics. For the statistical tests, the optimal value of p < 0.001 represents the significance threshold (α = 0.001).

	Metric	**Description**	**Definition (or R-function)**	**Optimal value**
Statistical tests	Welch's t-test	Variation on correlation where the two samples have unequal variances. Hypothesis is that two populations have equal means.	stats::t.test(…)	$p < 0.001$
	Kolmogorov-Smirnov test (*KS*)	Tests whether samples come from the same population, i.e. follow the same distribution.	stats::ks.test(…)	$p < 0.001$
	Cramér-von Mises (*CvM*)	Addresses limitations of KS test: (1) less focused on the central distribution; (2) more equal weighting on the tails of the distribution.	cramer::cramer.test(…) (Franz, 2014)	$p < 0.001$
Distribution	Hydrologic alteration factor (*HAF*)	A factor developed as part of the Indicators of Hydrologic Alteration (Mathews and Richter, 2007). Tests the replicability of sections of the probability distribution (lower-tail, IQR and upper-tail) for a given index.	$\frac{F_{sim} - F_{obs}}{F_{obs}}$ Where F is frequency, the no. values lying within the probability distribution.	0
Measures of error	Mean arctangent absolute percentage error (MAAPE)	A modification of MARE. Considers the relative error as an angle rather than a slope, reducing the bias of large errors.	$\frac{1}{n}\sum \arctan\left(\frac{I_{obs} - I_{sim}}{I_{obs}}\right)$ Where I is the index value and n the no. observations.	0
	Model efficiency (*NSE*)	Nash Sutcliffe efficiency. A measure of the goodness of fit of the HI to the 1:1 line.	$1 - \frac{\sum(I_{obs} - I_{sim})^2}{\sum(I_{obs} - \overline{I_{obs}})^2}$ Where I is the index value.	1

3 Results

3.1 Model parameters

For all catchments the low-flow optimised six-parameter GR6J model was invalidated; GR5J was invalidated for all catchments with the exception of the Tarland Burn and River Trent. A summary of the number of parameter sets (per model per catchment) and inter-quartile ranges is presented Table 4. For further details see Fig. B1. Being related in function, the parameters of the production (*x1*) and routing (*x3*) store capacities exhibit the greatest range. The groundwater exchange coefficient (*x4*) and inter-catchment exchange threshold (*x5*; where applicable) appear more consistent, whilst the time elapsed for the routing of flow appears inversely related to BFI.

Table 2. Interquartile (IQR) range across the parameter sets for each catchment.

	Tarland Burn		River Ribble	River Trent		River Nar	River Thrushel
No. free parameters	4	5	4	4	5	4	4
No. parameter sets	15	4	24	12	4	23	18
x1	322.8	837.3	533.1	49.1	90.9	342.2	492.9
x2	3.9	1.4	7.7	3.2	2.3	1.1	2.1
x3	154.7	244.9	176.5	71.6	497.1	298.7	169.4
x4	2.7	2.6	0.9	3.1	1.1	0.3	0.7
x5	-	1.4	-	-	2.4	-	-

3.2 Model evaluation

The ability of the covariance approach in the replication of the ER HIs is considered in terms of performance and consistency. The models are evaluated with reference to the metrics summarised in Table 3 previously. Results are considered by metric, with a focus on the ER HIs with the best and worst performance and consistency.

3.2.1 Statistical tests

A series of tests were applied with a view to determining if, statistically speaking, the observed and simulated ER HIs come from the same population. The tests focus on the mean (t-test), the central distribution (KS) and tails of the distribution (CVM test). Table B1 in the appendix details, per ER HI and catchment, the percentage of the parameter sets which did not show a significant level of agreement.

The statistical tests saw perfect agreement across all six timing indicators. With respect to the magnitude indices, the ER HI *BFIr* and the three skewness indicators do not satisfy any of the tests; performance appears irrespective of importance indicated by the hydroecological model or catchment. Magnitude median-quantile ratios agreement was mixed, with high and low flows achieving poor and good agreement respectively. Broadly, frequency indicators indicate a lack of agreement, with only the *PlsFld* index in the River Thrushel exhibiting performance and consistency. The role of statistical importance in the replication of these more complex indicators is also suggested, with *PlsQ75* replicated well in the Tarland (importance 0.69) and poorly in the Trent (importance 0.03). More broadly, log-transformed indicators saw better agreement; for example, the more important *MaxMonthlyVar* generally performed poorly, whilst *MaxMonthlyLogVar* saw agreement across all tests and parameter sets.

3.2.2 Distribution – Hydrologic alteration factor (HAF)

The hydrologic alteration factor (HAF) is a test of the replicability of the shape of the probability distribution. Fig. 3 summarises the HAF value across the central distribution and tails for each ER HI. There is agreement across the percentile ranges for the majority of the ER HIs considered. Notably, the 19 (of 22; statistics, log-ratios and quantile-median ratios) magnitude indicators not pictured achieved optimal HAF of zero. The three-monthly indicators (depicted) again highlight relative success in replicating a log-transformed index.

The performance of the six indicators capturing flow pulse events is varied: the central distribution of flood pulses is well-replicated whilst the upper tail exhibits a consistent large negative bias. The HAF values also serve to highlight some inconsistencies in the performance of the timing indicators. A variable negative bias is in evidence for the index *Mn7MaxJD*, however, in this case, it is worth noting that it is inherently more difficult for a hydrological model to detect and replicate (multiple) short-term events (Pool et al., 2017). Perhaps surprisingly, *Mn90MnJD* is subject to a large positive bias in the lower tail, i.e. the range of the distribution is underestimated. In contrast to *Mn7MaxJD*, this discrepancy may be due to the long(er)-term duration; with seasons of approximately 180 days in length, there are a limited number of values the indicator can take.

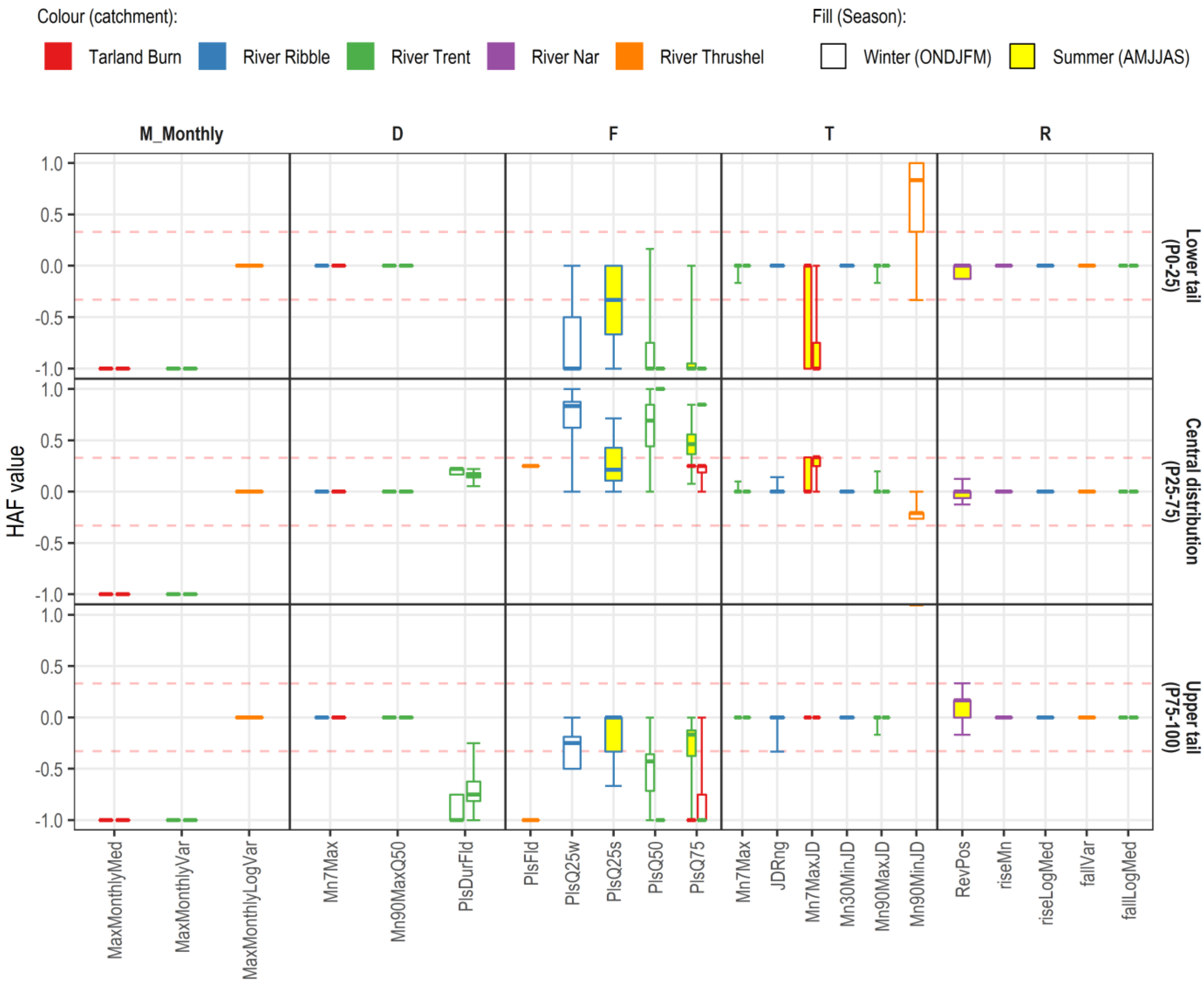


Figure 3. Hydrologic alteration factor (HAF) values for the three percentile ranges for each ER HI. The acceptable range of HAF values is defined as ±0.33 (red dashed line); HAF > 0 represents an increase in frequency relative to the observed whilst HAF < 0 represents a decrease. All magnitude statistic and ratio ER HIs achieved optimal values (HAF = 0) and are not depicted. The 4- and 5-parameter results are adjacent, left and right respectively, for the Tarland Burn and River Trent.

3.2.3 Error – MAAPE and NSE

Two measures of error were applied, MAAPE, a modification of the mean absolute relative error (MARE) which reduces the bias of large errors, as well as the more commonplace Nash Sutcliffe efficiency (NSE).

The MAAPE for each ER HI is depicted in Fig. 4; to ensure consistency with HAF, acceptable boundaries are specified as ±0.33 (depicted, horizontal red lines). Overall, the same general patterns may be observed; for example, skew indicators are not well replicated, log-transformation improves the monthly index performance, and timing, with the exception of

Mn90MinJD, achieves consistently good performance. However, it is clear that the consideration of multiple parameter sets per catchment model leads to variation in the simulated ER HI which may not have been detected by the previous metrics. MAAPE also serves to highlight the difference in performance across the median-quantile ratios, extreme high-flow indices (*Qmax* to *Q05*) are over-estimated whilst the replication of low-flow indices is subject to considerably less (negative) bias.

Figure 4. Mean arctangent absolute percentage error (MAAPE) values for each ER HI. As per HAF, the acceptable range is defined as ±0.33 (red dashed line). The 4- and 5-parameter results are adjacent, left and right respectively, for the Tarland Burn and River Trent.

The NSE is a measure of the efficiency or skill of the model relative to the observational mean (the 1:1 line; Murphy (1988); Gupta et al. (2009)), NSE < 0 indicates that the mean may be a better estimate. In Fig. 5, only ER HI with NSE > 0 are depicted with the number of parameter sets described as *n*; for all ER HI see Fig. B2.

Seventeen ER HI achieved NSE values greater than zero; further, the low values of n which are in evidence indicate a lack of consistency across parameter sets. Those ER HI which have already been shown to perform well are indicated, examples include the low flow median-quantile ratios, the log-transformed monthly index and the timing indicators more generally.

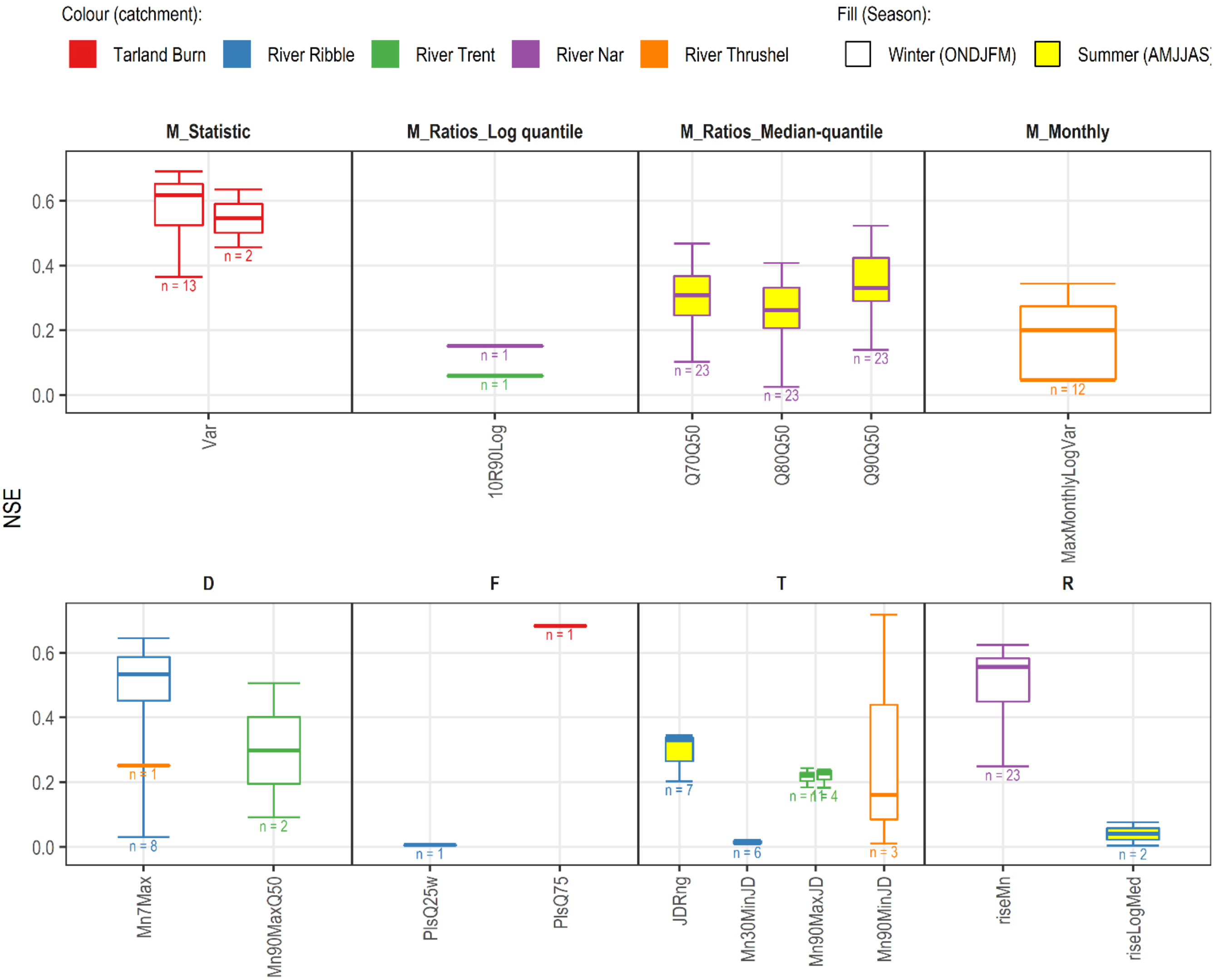


Figure 5. Nash Sutcliffe Efficiency (NSE) for each ER HI where NSE > 0 (model skill greater than observational mean); see Fig. B2 for all NSE. The 4- and 5-parameter results are adjacent, left and right respectively, for the Tarland Burn and River Trent.

3.2.5 Catchment level

In order to validate the modified covariance approach a range of catchments were considered; the main differences were in the BFI, length of the available time-series and the ER HIs considered. In this study, BFI ranged from 0.25 to 0.91, essentially flashy to groundwater-fed. With the exception of model parameterisation, there was no discernible difference in the replication of ER HIs. Similarly, the length of the available time-series appears to have made no observable difference to the replicability of the ER HI distributions specifically. In terms of error, MAAPE and NSE, lower overall performance for the shorter time-series is expected as a result of sample size sensitivity. Finally, despite consideration of a range of ER HIs with different associated importance, there appears a consistent message in terms of the performance and consistency of similar indices and the facets of the flow regime more broadly.

4 Discussion

There is a clear need to understand the impact of hydrologic change on the river ecosystem. To this end, hydrological models are used to simulate flow time-series from which ecologically relevant hydrological indicators are derived. Previous studies (e.g. Vis et al. (2015), Shrestha et al. (2014) and Pool et al. (2017)) have highlighted the inability of hydrological models to simulate a range, or suite, of ER HIs concurrently. In this study, a modification of Vogel and Sankarasubramanian (2003) covariance approach was applied to five hydrologically distinct catchments; the focus was on the replication of a suite of ER HIs identified through catchment-specific hydroecological models. The ability of this modified covariance approach was assessed through a series of evaluation metrics.

4.1 Performance and consistency

The consideration of a range of catchments provides a clear picture of the capacities of the hydrological models as well as the relative success of the covariance approach. Overall, replication of the ER HIs was good. Timing and log-transformed indicators (*logQVar*, *MaxMonthlyLogVar* and the log quantile ratios) were among the most consistent and well-replicated across the range of catchments, whilst difficulties were observed in replicating frequency and rate of change indices. Replication of indicators incorporating the seasonal median flow (Q50) was also poor, with large positive biases frequently observed. This may be observed directly through comparison of the replication of *Q01* and *Q01Q50* in the River Trent where the degree of error can be seen to markedly increase.

4.2 Advantages and limitations of the modified covariance approach

In this section we consider the general advantages of the modified covariance approach, over traditional calibration approaches, followed by the hydroecological modelling requirements. It is clear that no approach has been able to achieve adequate performance and consistency in the replication of more complex ER HIs, specifically those related to rate of change. Shrestha et al. (2014) observed difficulties in replicating low flows, the duration of flow pulses, and monthly flows specifically. In this study, no such observations have been made with regards to low flows and duration, indeed, these may be considered to be relatively well-replicated across all catchments. Poor replication of monthly ER HIs does however persist; log-transformed variations of these indicators may represent a viable alternative. Whilst Pool et al., 2017 saw improvements (relative to Shrestha et al. (2014) and Vis et al. (2015)), the need to calibrate the model to each ER HI in question would strongly call into question the reliability of the hydrological model (due to the inability of the hydrological model to simulate catchment hydrological processes simultaneously). The consistency with which (the majority of the) ER HIs are replicated here illustrates that this is not a necessary limitation of hydrological models. A lack of consistency in ER HIs demonstrating elevated levels of variability, such as high flows, is to be expected due to the dynamic nature of inter-annual weather patterns (Pool et al., 2017).

4.2.1 General advantages

Here follows a brief discussion of the general advantages of the modified covariance approach. First, uncertainty is reduced via a number of avenues:

- *Disinformative data:* Models calibrated following a traditional approach are particularly sensitive to measurement error (Westerberg et al., 2011). Lack of agreement in the observed-simulated time-series, even for a single event, may bias the objective function, leading to rejection of an otherwise well-performing parameter set (Beven, 2010; Westerberg et al., 2011). Methods which do not focus on the replication of time-series directly, such as the modified covariance approach, are known to limit the influence of input uncertainty (Westerberg et al., 2011; Euser et al., 2013).
- *Validation of model structure:* Consideration of the observed and simulated moments allows the user to validate the ability of the hydrological model structure in capturing the hydrological processes in the catchment, thus ensuring the selection of the optimal model.
- *Equifinality:* Equifinality, reaching the same outcome by different means, is a major challenge of hydrological modelling. In the modified covariance approach the entire parameter space is considered at the outset. A plausible parameter space is determined by focussing on the region which is best able to replicate the characteristics of the HIs, thereby reducing the uncertainty associated with equifinality (Wu et al., 2017).

Finally, whilst the large number of simulations required under the modified covariance approach may seem prohibitive, this demand may be offset. Unlike the calibration-validation paradigm, where selection algorithms may introduce issues of speed and accuracy (Seibert, 2000), finite time is needed to apply the covariance approach. All simulations of the hydrological model are performed at the outset; once the full suite of parameter sets have been simulated the hydrological model need not be run again. Under a more traditional approach, such as in (Pool et al., 2017) where the ER HIs serve as the objective, the HIs must be specified at the outset. This is not the case in the modified covariance approach, where the *n* Monte Carlo simulations can be performed in advance of HI selection. Thus, multiple suites of ER HIs may be considered (e.g. all rate of change or magnitude indicators) with limited additional time outlay.

4.2.1 Hydroecological model requirements

The explicit consideration of the outcomes of hydroecological modelling is perhaps both the most significant advantage and disadvantage of the modified covariance approach. Whilst hydrological modelling informed by the outcomes of hydroecological studies is not new, for instance, Pool et al. (2017) was informed by Knight et al. (2014), the novelty of this approach lies in the explicit consideration of the statistical importance of the ER HIs, identified through hydroecological modelling.

Following a traditional calibration approach, there is a practical limit to the number of ER HIs which may be considered; for instance, Pool et al. (2017)limited their multi-criteria objective function to five equally weighted indicators. However, by considering the relative importance of each ER HI, the modified covariance approach allows a large suite of ER HIs (seven to thirteen) to be considered with no apparent penalties. Further, contrary to expectations, a large number of important ER HIs (> 0.5) has no impact on replicability. In the case of the River Ribble, where a total of thirteen ER HIs were considered, seven had an importance greater than 0.5. Similarly, through this approach, a high weighting is not needlessly attributed to ER HIs with low importance.

The need for a hydroecological model represents the major limiting factor due to the requirement for long-term hydroecological time-series. Historically, hydrological and ecological data were collected for different objectives (Poff and Allan, 1995; Knight et al., 2008; Monk et al., 2008), leading to a mismatch in temporal and spatial coverage. High levels of disparity in sampling and gauging sites inevitably introduce noise into the model. However, the availability of national ecological datasets, such as BIOSYS in the UK , may serve to offset the issue of data availability. Such datasets may be used to develop regional hydroecological models based on flow regime type and the assumption of homogeneity in environmental conditions. The modified covariance approach may also be applied without a numerical measure of the relative importance of each indicator, this would however introduce an element of subjectivity into the parameterisation of the model.

4.3 General observations

4.3.1 Suitability of ER HIs in hydrological modelling

This, and previous studies, have observed difficulties in the replication of frequency ER HIs (flow pulses). This begs the question: Is this a product of the covariance approach? An inherent limitation of hydrological models more generally? Or is this related to the nature of the indicator itself? A review of the simulated flow suggests the latter. There is a tendency for the simulations to identify shorter more frequent pulses, whilst the observed pulses are longer and less frequent. For instance, the median error (MAAPE) for *PlsQ50* (the number of pulses above a baseline Q50 threshold) on the River Trent was 0.75; this falls to 0.368 if the focus is on the total duration of the pulses. The pooling of events with an inter-event time below some threshold, as per the inter-event time and volume criterion (Gustard and Demuth, 2009) for example, may serve to improve the replication of the pulse indicators. It should be noted that, in this study, this limitation does not extend to flood pulses (*FldPls*) due to the much larger inter-event time, thus allowing for better replication of flood pulses overall.

In multiple cases, this study observed difficulties in replicating those ER HIs which are considered relative to the median seasonal flow. Comparison of the indicators *Q01* and *Q01Q50* in the same catchment indicates that the lack of direct consideration of median flows in the parameterisation of the model may be a limiting factor. Indeed, it may be that the decomposition of such indicators into their component parts, e.g. Q01 and Q50, may lead to better replicability overall. Similarly, the results indicate that log-transformation of flows may lead to improvements in the replicability of certain ER HIs. Further work is required to confirm this premise.

4.3.2 Suitability of evaluation metrics

There is a lack of consistency in the evaluation metrics considered in the evaluation of hydrological model performance. Further, these studies make use of metrics which exhibit known bias, for example, mean absolute relative error (MARE; Kim and Kim (2016) and NSE (Gupta et al., 2009; Pushpalatha et al., 2012; Vis et al., 2015). For the measure of error, this study replaced the former with MAAPE (see Table 3). The reasons for the consideration of NSE in this study were twofold: (1) application of NSE is the norm; and (2) to illustrate the limitations of this measure. The limitations of NSE are frequently cited as low scores where there is high variability (Gupta et al., 2009) as well as a bias towards high flows (Pushpalatha et al., 2012). Additionally, the NSE is scaled by the standard deviation, rendering it incomparable across catchments (Gupta et al., 2009). In this study, only seventeen of the ER HIs achieved NSE > 1, i.e. the simulations are better than an estimation based on the observed mean. Similar observations were made in Vis et al. (2015). It can be concluded that, given this lack of robustness, NSE is not a suitable evaluation metric in studies such as this one.

4.4 Wider applicability and further work

The modified covariance approach is able to provide statistically robust simulations and projections of ER HIs for applications such as environmental flow assessment or in assessing the hydroecological impact of climate change as in Visser et al. (2018b) and Visser et al. (2019). However, the applicability of the approach is not limited to hydroecological studies and the simulation of ER HIs, being suited to the simulation of any HIs or hydrological signatures. Indeed, a focus on hydrological signatures may serve to improve the simulation of underlying hydrological processes more generally (Seibert, 2000; Euser et al., 2013). In this context, example applications include the replication of water resource management indicators (monthly, seasonal and annual flows). Such applications would require consideration of a statistical model for the determination of the statistical importance of indicators. The approach may also be used in the development of regional hydrological models, thereby facilitating the simulation of HIs in ungauged catchments. Finally, the clarity with which model structures are accepted or rejected makes the approach ideally suited for use in combination with model selection frameworks such as the Framework for Assessing the Realism of Model Structures (FARM; Euser et al. (2013)).

5 Concluding remarks

This study considered the performance and consistency of a modified covariance approach in the replication of ecologically relevant hydrological indicators. Application across five hydrologically diverse catchments showed a consistent level of performance across the majority of ER HIs; the timing facets of the flow regime were best replicated, whilst rate of change indicators saw the poorest performance and consistency. Relative to similar studies, there was an overall improvement in consistency, thus, this study represents an important advancement towards the robust application of hydrological models for ecological flow studies. The explicit consideration of hydroecological modelling outcomes allows the hydrological model to be tuned to parameters based on statistical importance. A further major advantage of the modified covariance approach lies in the identification of the plausible parameter space which best captures (replicates) the characteristics of the ER HIs, thereby providing a greater understanding of the suitability, limitations and uncertainties of the hydrological model structure.

Data availability: The hydroclimatological data used for all catchments (except the Tarland Burn) is freely available from the NRFA (2018), Met Office (2018a) and Met Office (2018b). Data for the Tarland Burn was provided to Heriot-Watt on request for this study by the James Hutton Institute (JHI, 2018).

Author contributions: AV developed the methodology, code and performed the data analysis. AV prepared the manuscript whilst LB provided review and edits. Both LB and SP provided supervision.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements: The authors gratefully acknowledge funding from the Engineering and Physical Science Research Council through award 1786424.

Appendix A – Method

A.1 Case studies

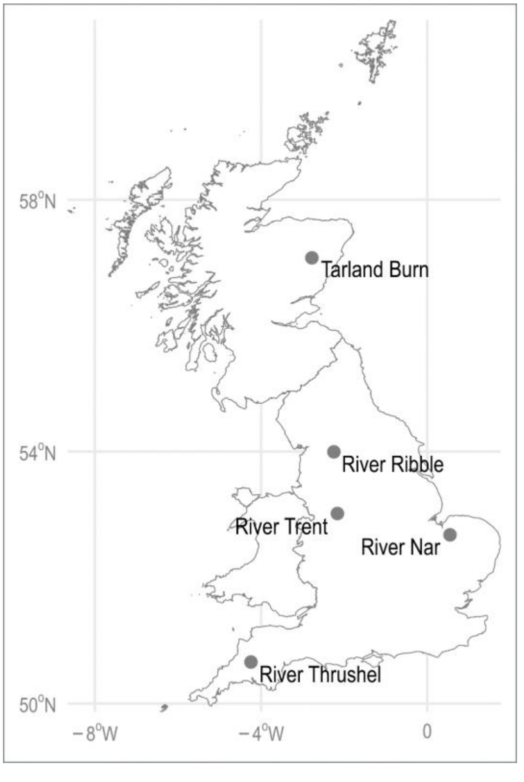


Figure A1. Distribution of the case study catchments across the UK.

A.2 Hydroecological modelling

Based on Olden and Poff (2003) and Monk et al. (2006), daily mean flow data was used to derive 63 hydrological indices per hydrological season: winter (ONDJFM) and summer (AMJJAS); for data source, see Table 1. Principal Component Analysis (PCA) was applied to identify those indices which describe the major aspects of the flow regime whilst minimising redundancy. Macroinvertebrates serve as the proxy for ecological response. Response is determined using the Lotic-Invertebrate Index for Flow Evaluation, accounting for macroinvertebrate flow velocity preferences (Extence et al., 1999). For four out of five case studies LIFE scores were determined to family level; data for the River Nar, obtained directly from the Environment Agency, was available to species level. The modelling focused on spring ecological activity (the period of peak activity and largest consistent availability of data).

After Visser et al. (2019), an Information Theory approach to modelling was taken in order to provide a quantitative measure of support for parameters and candidate models. Inference is made from multiple models through model averaging. In summary: (1) the candidate models are evaluated with respect to the second-order bias corrected Akaike Information Criterion (AICc) (after Burnham and Anderson (2002); see also Visser et al. (2019)); (2) a best approximating model is inferred from a weighted combination of all the candidate models; (3) the parameters are ranked, such that the highest value represents the most important in the model; (4) filters are applied to remove parameters where the estimate and confidence intervals are zero (i.e. certainty that the index is not to be included) and to reduce the model to the parameters which describe 95% of the cumulative information. For further details, see Visser et al. (2018a) and Visser et al. (2019).

A.3 Hydrological modelling

Table A1. Parameter limits for the hydrological models.

	Description	Limits
x1	Capacity of production store (mm)	(100,1200}
x2	Groundwater transfer (mm/day; positive indicates flow *from* aquifer)	(-5,25}
x3	Capacity of routing store (mm)	(20,1000}
x4	Time lag between rainfall event and flow (days)	(0.5,30}
x5	Inter-catchment exchange threshold (-)	(-5,25}
x6	Capacity of parallel routing store (mm)	(20,1000}

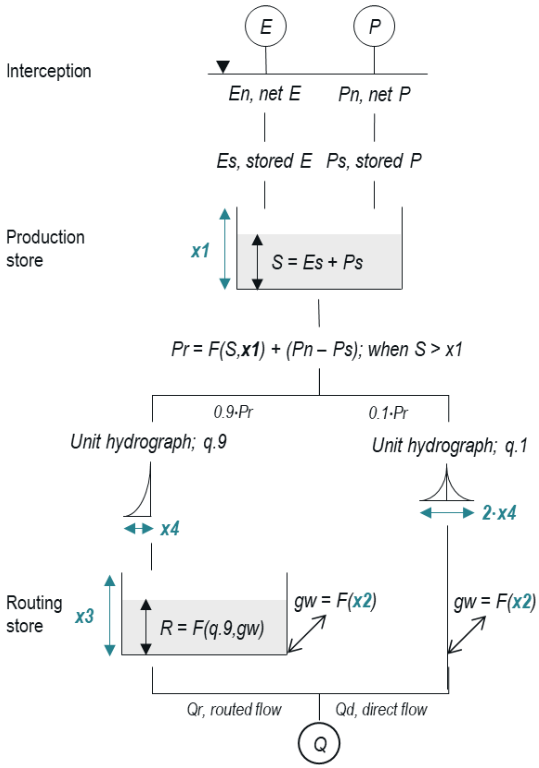


Figure A2. Structure of the GR4J hydrological model; based on Perrin et al. (2003). The 5-parameter GR5J sees the addition of x5, inter-catchment exchange parameter, at the same locations as x2, whilst GR6J sees the addition of a store parallel, capacity x6, to the routing store.

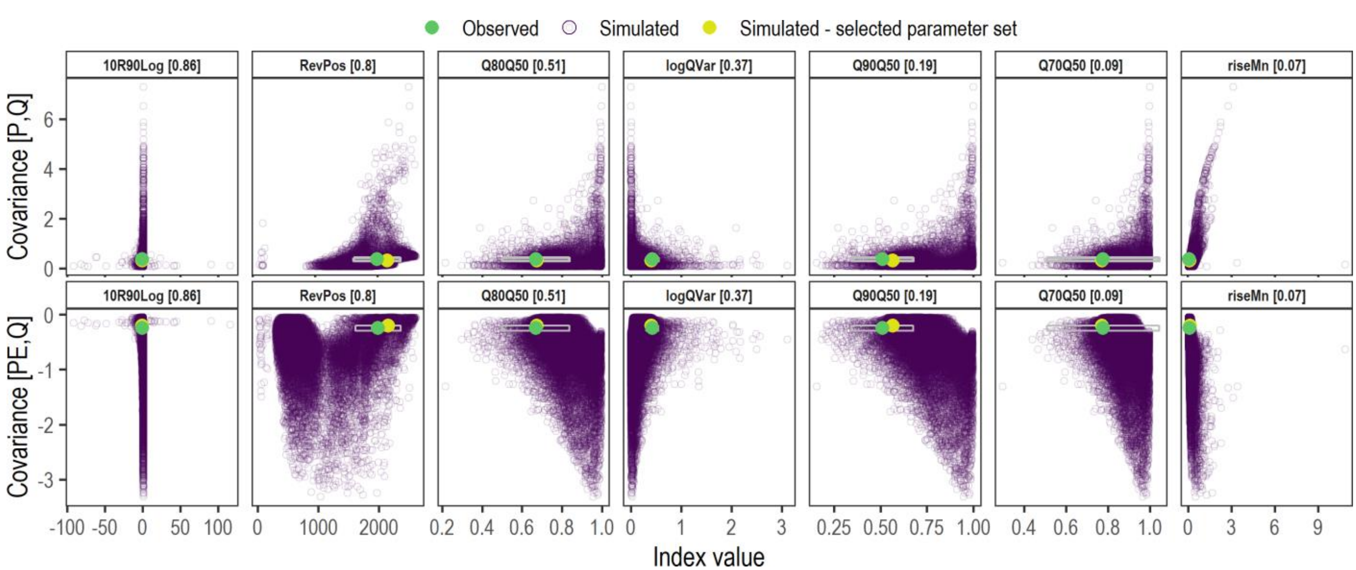


Figure A3. Observed and simulated moments for the 100,000 Monte Carlo simulations using the GR4J model for the River Nar case study. The grey boxes depict the boundaries of the limits of acceptability per index. One of the selected parameter sets, *i = 73,952*, is highlighted (yellow).

Appendix B – Supplementary results

B.1 Ecologically relevant hydrological indices and test statistics

Table B1. Ecologically relevant hydrological index descriptions; grouping is by facet of the flow regime. Seasons are indicated through no shading (winter) and shading (summer). Subsequent columns are catchment specific, denoting ER HI importance, and the results of the statistical tests detailed in Table 3. In the table, a flood threshold is the flow equivalent for a flood recurrence interval of 1.67 years (on the baseline).

***Four and five parameter models were applied to both the Tarland Burn and River Trent. Single digit entries should be interpreted as being the same across both models; where entries are separated, e.g. for *10R90Log*, the former represents GR4J and the latter GR5J.**

			Tarland Burn*					River Ribble					River Trent*					River Nar					River Thrushel				
Index	**Description**	**Units**	**Importance**	**Normal**	**Mean**	**IQR**	**Tails**	**Importance**	**Normal**	**Mean**	**IQR**	**Tails**	**Importance**	**Normal**	**Mean**	**IQR**	**Tails**	**Importance**	**Normal**	**Mean**	**IQR**	**Tails**	**Importance**	**Normal**	**Mean**	**IQR**	**Tails**
Magnitude - Statistic																											
IQR	Interquartile range of flow.	m^3s^{-1}	0.43	0	0	0	0																				
Var	Variance in flow.	-	0.46	0	0	0	0																				
Q01	Q1 flow (extreme high flow).	m^3s^{-1}						0.32	0	83.3	70.8	87.5															
Sk	Skewness, mean relative to median.	-						0.11	0	100	100	100											0.88	0	100	100	100
SkRel	Relative skewness, mean minus median, relative to median.	-						0.09	0	100	100	100															
Sk100	Range relative to the median.	-																					0.38	0	100	100	100
logQVar	Variance in log-transformed flow.	-																0.37	0	0	0	0					
BFIr	The seasonal BFI relative to baseline BFI.	-											0.03	0	100	100	100										

Magnitude - Ratios - Log quantile																											
10R90 Log													0.97	100	0	8.3 — 100	0	0.86	100	0	0	0					
20R80 Log	Log-transformed ratio, xx^{th} to yy^{th} percentile flow.	-																					0.94	100	0	0	0
25R75 Log													0.97	100	0	8.3 — 100	0										
Magnitude - Ratios - Median-quantile																											
MaxQ50	Max. flow relative to median (extreme high flow).	-																					0.32	0	100	100	100
Q01Q50								0.53	0	100	100	100	0.03	0	100	100	100										
Q05Q50	Q*xx* flow relative to median (high flow).	-						0.4	0	100	100	100															
Q20Q50													0.03	0	16.7 — 100	66.7 — 100	58.3 — 100										
Q60Q50																							0.97	0	72.2	33.3	72.2
Q70Q50	Q*xx* flow relative to median (low flow).	-						0.88	0	83.3	66.7	83.3						0.09	0	0	0	0	0.99	0	77.8	55.6	72.2
Q80Q50								0.38	0	83.3	75	87.5						0.51	0	0	0	0					
Q90Q50																		0.19	0	0	0	0					

Magnitude - Monthly																											
Max Monthly Med	Median of max. monthly flow.	m^3s^{-1}	0.7	0	0—25	0	0																				
Max Monthly Var	Variability in max. monthly flow.	-	0.45	0	0	0	0						0.92	100	33.3—100	91.7—100	91.7—100										
Max Monthly LogVar	Variability in max. monthly log-transformed flow.	-																					0.45	100	0	0	0
Duration																											
Mn7 Max	Mean of the 7-day cumulative max. flow.	m^3s^{-1}	0.53	0	0	0	0	0.14	0	20.8	33.3	25											0.5	0	94.4	77.8	88.9
Mn90 MaxQ50	Mean of the 90-day cumulative max. flow relative to the median.	-	0.53	0	0	0	0						0.06	0	25—100	16.7—100	33.3—100										
PlsDur Fld	Duration of pulses above a (baseline) flood threshold.	Days											0.02	100	0	0	50—25										
PlsDur Q75Var	Variation in the duration of pulses below a Q75 (baseline) threshold.	-																									

Frequency																											
PlsFld	No. of pulses above a (baseline) flood threshold.	Count																					0.41	100	0	0	55.6
PlsQ25w	No. of pulses above a Q*xx* (baseline) threshold.	Count						0.64	0	91.7	83.3	95.8															
PlsQ25s								0.58	0	66.7	70.8	83.3															
PlsQ50													0.04	0	100	100	100										
PlsQ75	No. of pulses below a Q*xx* (baseline) threshold.	Count	0.69	0	0	0	0						0.03	0	100	100	100										
Timing																											
JDRng	Range in the Julian days for the max. and min. daily mean flow.	JD						0.73	0	0	0	0															
Mn7 MaxJD	Julian day of the mean 7-day max. flow.	JD	0.78	0	0	0	0						0.94	0	0	0	0										
Mn30 MinJD	Julian day of the mean 30-day min. flow.	JD						0.67	0	0	0	0															
Mn90 MaxJD	Julian day of the mean 90-day max. flow.	JD											0.03	100	0	0	0										
Mn90 MinJD	Julian day of the mean 90-day min. flow.	JD																					0.88	100	0	0	0

Rate of change																											
RevPos	No. days when flow increases (positive reversals).	Days																0.8	0	100	100	100					
riseMn	Mean rise rate (flow increasing).	m^3s^{-1}																0.07	0	0	0	0					
riseLog Med	Median log-transformed rise rate (flow increasing).	m^3s^{-1}						0.55	0	0	0	4.17															
fallVar	Variation in fall rate (flow decreasing).	-																					0.16	0	100	100	100
fallLog Med	Median log-transformed fall rate (flow decreasing).	m^3s^{-1}											0.93	100	91.7—75	100											

B.2 Model parameters

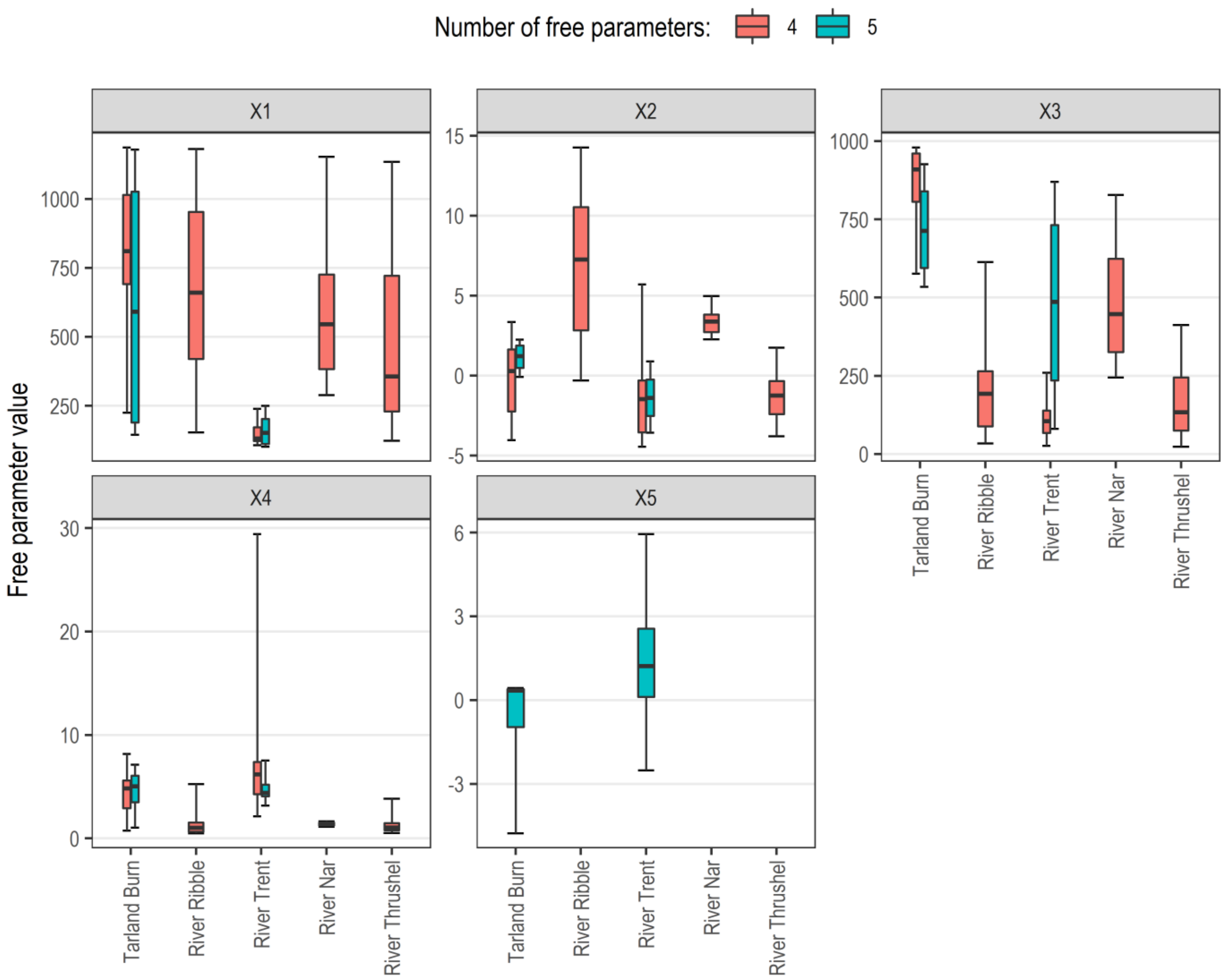


Figure B1. Boxplots of the parameter values across the 100 selected models. The whiskers represent the maximum and minimum values observed.

B.3 Nash Sutcliffe Efficiency

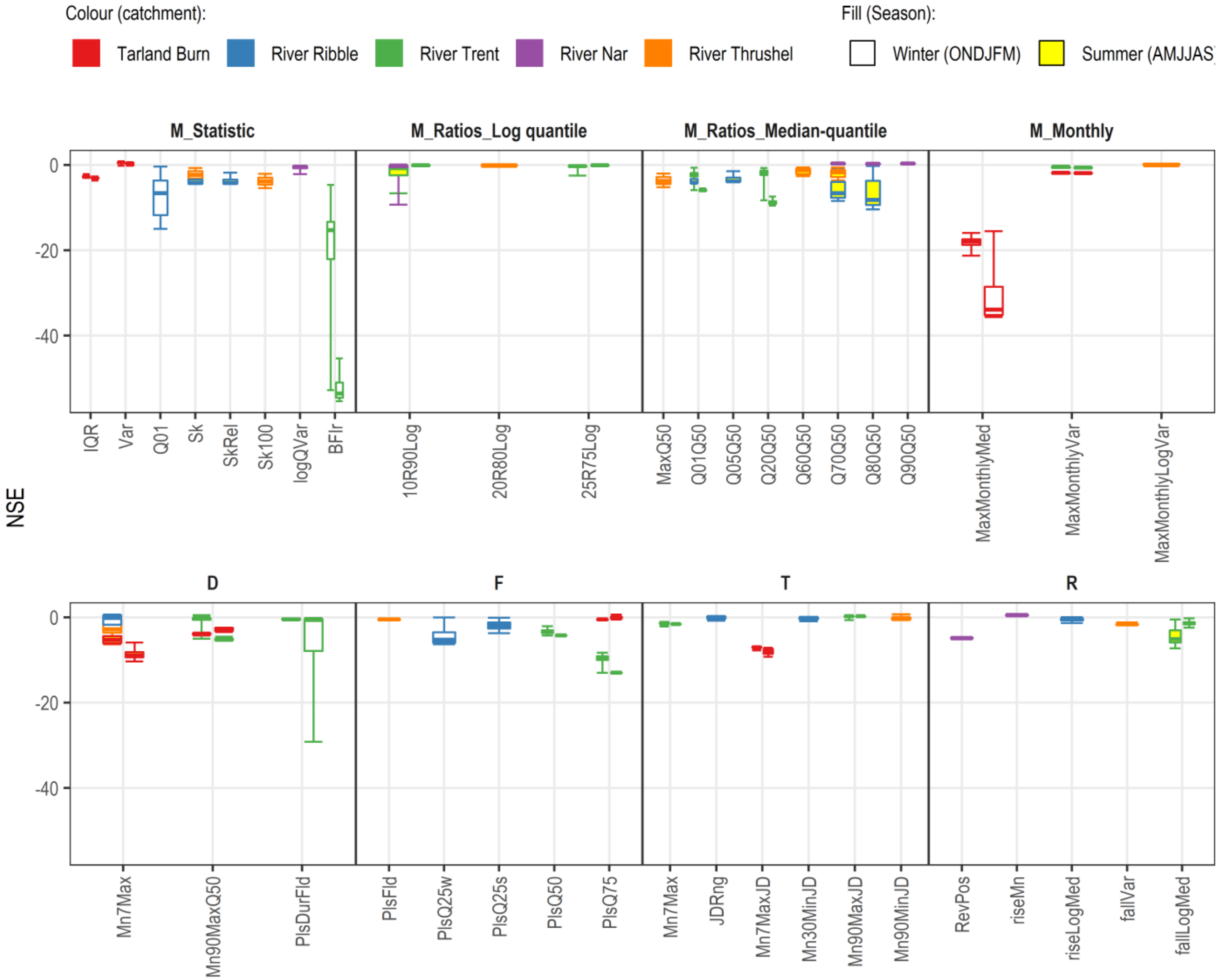


Figure B2. Nash Sutcliffe Efficiency (NSE) for each ER HI; see Fig. 5 for NSE > 0. The 4- and 5-parameter results are adjacent, left and right respectively, for the Tarland Burn and River Trent.

References

Anderson Michael, L., Chen, Z. Q., and Kavvas, M. L.: Modeling Low Flows on the Cosumnes River, Journal of Hydrologic Engineering, 9, 126-134, 10.1061/(ASCE)1084-0699(2004)9:2(126), 2004.

Anderson, T. W.: On the Distribution of the Two-Sample Cramer-von Mises Criterion, Ann. Math. Statist., 33, 1148-1159, http://dx.doi.org/10.1214/aoms/1177704477, 1962.

Andreassian, V., Bergstrom, S., Chahinian, N., Duan, Q., Gusev, Y., Littlewood, I., Mathevet, T., Michel, C., Montanari, A., and Moretti, G.: Catalogue of the models used in MOPEX 2004/2005, IAHS publication, 41 pp., 2006.

Arthington, A. H., Bunn, S. E., Poff, N. L., and Naiman, R. J.: The Challenge of Providing Environmental Flow Rules to Sustain River Ecosystems, Ecological Applications, 16, 1311-1318, http://dx.doi.org/10.1890/1051-0761(2006)016[1311:TCOPEF]2.0.CO;2, 2006.

Arthington, A. H.: Chapter 1 - River Values and Threats, in: Environmental Flows: Saving Rivers in the Third Millenium, edited by: Arthington, A., University of California Press, California, 2012.

Beven, K.: A manifesto for the equifinality thesis, Journal of Hydrology, 320, 18-36, http://dx.doi.org/10.1016/j.jhydrol.2005.07.007, 2006.

Beven, K. J.: Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models, Hydrological Processes, 24, 1537-1547, http://dx.doi.org/10.1002/hyp.7718, 2010.

Beven, K. J.: 1 Down to Basics: Runoff Processes and the Modelling Process, in: Rainfall-runoff modelling: The Primer, 2nd ed., Wiley-Blackwell, Chichester, 2012.

Blöschl, G., and Montanari, A.: Climate change impacts—throwing the dice?, Hydrological Processes, 24, 374-381, doi:10.1002/hyp.7574, 2010.

The Brisbane Declaration - Environmental Flows are Essential for Freshwater Ecosystem Health and Human Well-Being: https://www.conservationgateway.org/ConservationPractices/Freshwater/EnvironmentalFlows/MethodsandTools/ELOHA/Documents/Brisbane-Declaration-English.pdf, access: 23/08/2016, 2007.

Bunn, S. E., and Arthington, A. H.: Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity, Environmental Management, 30, 492-507, 2002.

Burnham, K. P., and Anderson, D.: Model Selection and Multi-model Inference: A Pratical Information-Theoretic Approch, Springer, New York, 488 pp., 2002.

Carlisle, M., Falcone, J., Wolock, M., Meador, R., and Norris, H.: Predicting the natural flow regime: models for assessing hydrological alteration in streams, River Research and Applications, 26, 118-136, http://dx.doi.org/10.1002/rra.1247, 2010.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resources Research, 48, W05552, http://dx.doi.org/10.1029/2011WR011721, 2012.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V.: The suite of lumped GR hydrological models in an R package, Environmental Modelling & Software, 94, 166-171, 10.1016/j.envsoft.2017.05.002, 2017.

airGR: Suite of GR hydrological models for precipitation-runoff modelling. R package version 1.0.15.2: http://webgr.irstea.fr/modeles/?lang=en, 2018.

Cramér, H.: On the composition of elementary errors, Scandinavian Actuarial Journal, 1928, 13-74, https://doi.org/10.1080/03461238.1928.10416862, 1928.

Davis, J., O'Grady, A. P., Dale, A., Arthington, A. H., Gell, P. A., Driver, P. D., Bond, N., Casanova, M., Finlayson, M., Watts, R. J., Capon, S. J., Nagelkerken, I., Tingley, R., Fry, B., Page, T. J., and Specht, A.: When trends intersect: The challenge of protecting freshwater ecosystems under multiple land use and hydrological intensification scenarios, Science of The Total Environment, 534, 65-78, 10.1016/j.scitotenv.2015.03.127, 2015.

Efstratiadis, A., and Koutsoyiannis, D.: One decade of multi-objective calibration approaches in hydrological modelling: a review, Hydrological Sciences Journal, 55, 58-78, 2010.

Freshwater and Marine Biological Surveys for Invertebrates England (BIOSYS): https://data.gov.uk/dataset/ae610ec8-7635-4359-9662-c920046950f7/freshwater-and-marine-biological-surveys-for-invertebrates-england, access: 10/10/2018, 2018.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst. Sci., 17, 1893-1912, doi.org/10.5194/hess-17-1893-2013, 2013.

Extence, C. A., Balbi, D. M., and Chadd, R. P.: River flow indexing using British benthic macroinvertebrates: a framework for setting hydroecological objectives, Regulated Rivers: Research & Management, 15, 545-574, 10.1002/(sici)1099-1646(199911/12)15:6<545::aid-rrr561>3.0.co;2-w, 1999.

cramer: Multivariate nonparametric Cramer-Test for the two-sample-problem. R package version 0.9-1. https://cran.r-project.org/web/packages/cramer/index.html, 2014.

Gleick, P. H.: Water in crisis: Paths to sustainable water use, Ecological Applications, 8, 571-579, 10.1890/1051-0761(1998)008[0571:WICPTS]2.0.CO;2, 1998.

Gleick, P. H.: Water strategies for the next administration, Science, 354, 555-556, 2016.

Grayson, R., and Blöschl, G.: Summary of pattern comparison and concluding remarks, in: Spatial Patterns in Catchment Hydrology: Observations and Modelling, edited by: Grayson, R., and Blöschl, G., Cambridge University Press, Cambridge, UK, 355–367, 2001.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80-91, http://dx.doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrol. Earth Syst. Sci., 18, 463-477, 10.5194/hess-18-463-2014, 2014.

Gustard, A., and Demuth, S.: Manual on Low-flow Estimation and Prediction, Operational Hydrology Report No. 50, Publications Board World Meteorological Organization (WMO), Geneva, Switzerland, 136, 2009.

Haines, K., Blower, J. D., Drecourt, J. P., Liu, C., Vidard, A., Astin, I., and Zhou, X.: Salinity Assimilation Using S(T): Covariance Relationships, Monthly Weather Review, 134, 759-771, 10.1175/MWR3089.1, 2006.

Hughes, D. A.: Incorporating groundwater recharge and discharge functions into an existing monthly rainfall–runoff model/Incorporation de fonctions de recharge et de vidange superficielle de nappes au sein d’un modèle pluie-débit mensuel existant, Hydrological Sciences Journal, 49, null-311, 10.1623/hysj.49.2.297.34834, 2004.

Tarland Burn monitoring data. Received 11/10/2018. Available upon request., 2018.

Kim, S., and Kim, H.: A new metric of absolute percentage error for intermittent demand forecasts, International Journal of Forecasting, 32, 669-679, https://doi.org/10.1016/j.ijforecast.2015.12.003, 2016.

Klaar, M. J., Dunbar, M. J., Warren, M., and Soley, R.: Developing hydroecological models to inform environmental flow standards: a case study from England, Wiley Interdisciplinary Reviews: Water, 1, 207-217, 10.1002/wat2.1012, 2014.

Knight, R. R., Brian Gregory, M., and Wales, A. K.: Relating streamflow characteristics to specialized insectivores in the Tennessee River Valley: a regional approach, Ecohydrology, 1, 394-407, 2008.

Knight, R. R., Gain, W. S., and Wolfe, W. J.: Modelling ecological flow regime: an example from the Tennessee and Cumberland River basins, Ecohydrology, 5, 613-627, 10.1002/eco.246, 2011.

Knight, R. R., Murphy, J. C., Wolfe, W. J., Saylor, C. F., and Wales, A. K.: Ecological limit functions relating fish community response to hydrologic departures of the ecological flow regime in the Tennessee River basin, United States, Ecohydrology, 7, 1262-1280, 10.1002/eco.1460, 2014.

Kokkonen, T. S., and Jakeman, A. J.: Chapter 14 Structural effects of landscape and land use on streamflow response, in: Developments in Environmental Modelling, edited by: Beck, M. B., Elsevier, 303-321, 2002.

Le Moine, N.: Le bassin versant de surface vu par le souterrain: une voie d’amélioration des performance et du réalisme des modéles pluie-débit? [French and English (in part)], Université Pierre et Marie Curie et Le Centre d'Antony, IRSTEA / Pierre and Marie Curie University and the Antony Centre, IRSTEA. Available from: http://www.sisyphe.upmc.fr/~lemoine/docs/These_Le_Moine_finale.pdf., 324 pp., 2008.

Lytle, D. A., and Poff, N. L.: Adaptation to natural flow regimes, Trends in Ecology & Evolution, 19, 94-100, 10.1016/j.tree.2003.10.002, 2004.

Mathews, R., and Richter, B.: Application of the Indicators of Hydrologic Alteration Software in Environmental Flow Setting1, JAWRA Journal of the American Water Resources Association, 43, 1400-1413, https://doi.org/10.1111/j.1752-1688.2007.00099.x, 2007.

McManamay, R. A., Orth, D. J., Kauffman, J., and Davis, M. M.: A Database and Meta-Analysis of Ecological Responses to Stream Flow in the South Atlantic Region, Southeastern Naturalist, 12, 1-36, 10.1656/058.012.m501, 2013.

MIDAS: UK Daily Rainfall Data: http://catalogue.ceda.ac.uk/uuid/c732716511d3442f05cdeccbe99b8f90, access: 11/10/2018, 2018a.

MIDAS: UK Hourly Weather Observation Data: http://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c, access: 11/10/2018, 2018b.

Monk, W. A., Wood, P. J., Hannah, D. M., Wilson, D. A., Extence, C. A., and Chadd, R. P.: Flow variability and macroinvertebrate community response within riverine systems, River Research and Applications, 22, 595-615, http://dx.doi.org/10.1002/rra.933, 2006.

Monk, W. A., Wood, P. J., Hannah, D. M., and Wilson, D. A.: Macroinvertebrate community response to inter-annual and regional river flow regime dynamics, River Research and Applications, 24, 988-1001, 10.1002/rra.1120, 2008.

Murphy, A. H.: Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient, Monthly Weather Review, 116, 2417-2424, 10.1175/1520-0493(1988)116<2417:ssbotm>2.0.co;2, 1988.

Murphy, J. C., Knight, R. R., Wolfe, W. J., and W, S. G.: Predicting Ecological Flow Regime at Ungauged Sites: A Comparison of Methods, River Research and Applications, 29, 660-669, https://doi.org/10.1002/rra.2570, 2012.

National Hydrological Monitoring Programme: https://nrfa.ceh.ac.uk/nhmp, access: 12/10/2018, 2018.

Olden, J. D., and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River Research and Applications, 19, 101-121, http://dx.doi.org/10.1002/rra.700, 2003.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of Hydrology, 303, 290-306, http://dx.doi.org/10.1016/j.jhydrol.2004.08.026, 2005.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275-289, 10.1016/S0022-1694(03)00225-7, 2003.

Perrin, C., Andréassian, V., Rojas Serna, C., Mathevet, T., and Le Moine, N.: Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments, Water Resources Research, 44, W08447, http://dx.doi.org/10.1029/2007WR006579, 2008.

Peters, D. L., Baird, D. J., Monk, W. A., and Armanini, D. G.: Establishing standards and assessment criteria for ecological instream flow needs in agricultural regions of Canada, Journal of environmental quality, 41, 41-51, 10.2134/jeq2011.0094, 2012.

Poff, N. L., and Allan, J. D.: Functional Organization of Stream Fish Assemblages in Relation to Hydrological Variability, Ecology, 76, 606-627, 10.2307/1941217, 1995.

Poff, N. L., Richter, B. D., Arthington, A. H., Bunn, S. E., Naiman, R. J., Kendy, E., Acreman, M., Apse, C., Bledsoe, B. P., Freeman, M. C., Henriksen, J., Jacobson, R. B., Kennen, J. G., Merritt, D. M., O’Keeffe, J. H., Olden, J. D., Rogers, K., Tharme, R. E., and Warner, A.: The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards, Freshwater Biology, 55, 147-170, 10.1111/j.1365-2427.2009.02204.x, 2010.

Pool, S., Vis, M. J. P., Knight, R. R., and Seibert, J.: Streamflow characteristics from modeled runoff time series – importance of calibration criteria selection, Hydrol. Earth Syst. Sci., 21, 5443-5457, https://doi.org/10.5194/hess-21-5443-2017, 2017.

Power, M. E., Sun, A., Parker, G., Dietrich, W. E., and Wootton, J. T.: Hydraulic Food-Chain ModelsAn approach to the study of food-web dynamics in large rivers, BioScience, 45, 159-167, http://dx.doi.org/10.2307/1312555, 1995.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V.: A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, Journal of Hydrology, 411, 66-76, http://dx.doi.org/10.1016/j.jhydrol.2011.09.034, 2011.

Pushpalatha, R., Perrin, C., Moine, N. L., and Andréassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, Journal of Hydrology, 420-421, 171-182, https://doi.org/10.1016/j.jhydrol.2011.11.055, 2012.

Richter, B. D., Baumgartner, J. V., Powell, J., and Braun, D. P.: A Method for Assessing Hydrologic Alteration within Ecosystems, Conservation Biology, 10, 1163-1174, 10.1046/j.1523-1739.1996.10041163.x, 1996.

Rivers Task Team: Reference Condition Descriptions for Rivers in Great Britain, UK TAG, United Kingdom, 2004.

Seibert, J.: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, Hydrol. Earth Syst. Sci., 4, 215-224, 10.5194/hess-4-215-2000, 2000.

Shrestha, R. R., Peters, D. L., and Schnorbus, M. A.: Evaluating the ability of a hydrologic model to replicate hydro-ecologically relevant indicators, Hydrological Processes, 28, 4294-4310, https://doi.org/10.1002/hyp.9997, 2014.

Smith, M. B., Koren, V., Reed, S., Zhang, Z., Zhang, Y., Moreda, F., Cui, Z., Mizukami, N., Anderson, E. A., and Cosgrove, B. A.: The distributed model intercomparison project – Phase 2: Motivation and design of the Oklahoma experiments, Journal of Hydrology, 418-419, 3-16, https://doi.org/10.1016/j.jhydrol.2011.08.055, 2012.

Tharme, R. E.: A global perspective on environmental flow assessment: emerging trends in the development and application of environmental flow methodologies for rivers, River Research and Applications, 19, 397-441, 10.1002/rra.736, 2003.

EflowStats: An R package to compute ecologically-relevant streamflow statistics: https://github.com/USGS-R/EflowStats, 2013.

Tong, C., and Graziani, F.: A Practical Global Sensitivity Analysis Methodology for Multi-Physics Applications, in: Computational Methods in Transport: Verification and Validation, Berlin, Heidelberg, 2008, 277-299,

Vis, M., Knight, R., Pool, S., Wolfe, W., and Seibert, J.: Model calibration criteria for estimating ecological flow characteristics, Water, 7, 2358-2381, https://doi.org/10.3390/w7052358, 2015.

Visser, A., Beevers, L., and Patidar, S.: Complexity in hydroecological modelling, a comparison of stepwise selection and information theory, River Research and Applications, 10.1002/rra.3328, 2018a.

Visser, A., Beevers, L., and Patidar, S.: The Impact of Climate Change on Hydroecological Response in Chalk Streams. , Preprints 2018, 10.20944/preprints201812.0266.v1, 2018b.

Visser, A. G., Beevers, L., and Patidar, S.: A coupled modelling framework to assess the hydroecological impact of climate change, Environmental Modelling & Software, https://doi.org/10.1016/j.envsoft.2019.01.004, 2019.

Vogel, R. M., and Sankarasubramanian, A.: Validation of a watershed model without calibration, Water Resources Research, 39, http://dx.doi.org/10.1029/2002WR001940, 2003.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., and Davies, P. M.: Global threats to human water security and river biodiversity, Nature, 467, 555, 10.1038/nature09440, 2010.

Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C. Y.: Calibration of hydrological models using flow-duration curves, Hydrol. Earth Syst. Sci., 15, 2205-2227, https://doi.org/10.5194/hess-15-2205-2011, 2011.

Wu, Q., Liu, S., Cai, Y., Li, X., and Jiang, Y.: Improvement of hydrological model calibration by selecting multiple parameter ranges, Hydrol. Earth Syst. Sci., 21, 393-407, https://doi.org/10.5194/hess-21-393-2017, 2017.

Wu, X., Heflin, M. B., Schotman, H., Vermeersen, B. L. A., Dong, D., Gross, R. S., Ivins, E. R., Moore, A. W., and Owen, S. E.: Simultaneous estimation of global present-day water transport and glacial isostatic adjustment, Nature Geoscience, 3, 642, 10.1038/ngeo938, 2010.