Review of the manuscript „Replication of ecologically relevant hydrological indicators following a modified covariance approach to hydrological model parameterisation“ by Visser et al.

Dear editor and authors,

I thank the authors for their well-structured answers to the comments. I especially appreciate that some major concerns have been addressed and are now improved in the current version of the manuscript. The use of 5 catchments instead of 1 catchment, the use of multiple hydrological models, and the use of 20 parameter sets instead of 1 parameter set strongly improved the quality of the work and clearly increased the relevance of the study for the ecohydrological community. The presentation of the results was also changed and results are now presented in a logically structured way. However, I still have some major concerns regarding some fundamental statements made in this manuscript about parameterization, objective functions, and split-sample tests.

Major comments

- It is multiple times highlighted that the proposed covariance approach has some considerable benefits over a traditional calibration approach. However, it is not clear what a traditional calibration approach is. A precise description of a traditional calibration approach is in my opinion highly important to be able to make a fair comparison between the two approaches.
- The covariance approach is described as an approach where model structure is validated (step 2 of approach) prior to parameterisation (step 3 of approach). I have two concerns regarding this statement. First, step 1 consists of sampling the parameter space and running simulations. So strictly speaking, simulations are run without validating the structure. To me a validation of the model structure without parameterization would e.g. be the check if the model structure represents the most important processes occurring in a catchment. For this no parameters are needed. Second, stage 2 of the approach is the visualization of the simulations in the form of efficiency plots, and stage 3 is the selection of parameter sets using covariance and ER His. Stage 2 is therefore not a validation – validation is only done in step 3 at the moment where a “judgment” is made about the validity of a parameter set.
- It is argued that the use of the covariance approach is a shift away from the use of objective functions. I don’t agree with this statement. Objective functions are criteria used to objectively select behavioural parameter sets. In this study, covariance and ecologically relevant hydrological indicators (ER HI) are used to select behavioural parameter sets. So both the covariance and the ER His are objective functions. Given that parameter sets are selected based on a combination of covariance and ER HIs, I would argue that a “classical” multi-objective function is used in this study. (The authors referenced in their response to the comments the review of Efstratiadis and Koutsoyiannis, where it is acknowleged that the term multi-objective can refer to both a scalar of multiple criteria or a vector based optimization approach). I am still not so happy with the term covariance approach, because the selection of parameter sets is based on two covariance metrics and many more ER HIs. The approach is therefore named after one out of many metrics used.
- The statement that the proposed covariance approach “avoids the need for split-sampling” is highly critical. The idea of split-sampling is that one selects parameter sets in one time period and then tests these selected parameter sets in an independent (maybe climatically/

meteorologically different) time period. This safety-check is independent of the approach used to select the parameter sets.

- In the discussion (chap. 4.4) the covariance approach is suggested as a tool to be used in environmental flow assessment, climate change studies, water resources management, prediction in ungauged catchments, and model selection frameworks. I wonder if the approach is really a solution to many of the biggest challenges in hydrology?
- The results of this study are mostly compared against three other studies that model ER HIs (i.e., Pool et al., 2017; Vis et al., 2015, Shrestha et al., 2014). Could you maybe find more studies for comparison? There are quite a number of studies using hydrological signatures (not necessarily ecologically relevant) for model calibration or validation. Such studies could be helpful to discuss the covariance approach and the simulation results of this study.

Minor comments

- P18 L20 - Study areas: Could you maybe provide some information on how exactly you selected the 5 study catchments?
- P18 L28 and Table 1: To complete the information in Table 1 I would add the information about the altitude, especially since you explicitly mention that altitude is different between the 5 study catchments. I would also suggest to either provide the definition of BFI in the table caption or not to use the abbreviation. The * behind “principle land use” should probably also be added to the table caption. Finally, you could eventually add the river type to the table.
- P19 L8: You mention that Perrin et al. (2008), Coron et al. (2016), and others have been using the GRJ model series for a wide range of applications. To increase the information content of this sentence, I recommend to explicitly state what GRJ has been used for (e.g. GR4 was used for climate change modelling, the prediction in ungauged basins, etc.).
- P19 L10-20: Parameter x3 is missing in the description of the model parameters. Given that the focus of the study is not the development of a new model I wonder if such a detailed description is needed. I suggest keeping the model description more generic focusing on the runoff processes represented by the model.
- P20 chpt. 2.4: The covariance approach is closely related to the GLUE methodology proposed by Beven and Binley (1992). Both approaches generate in a first step a large set of parameters, from which a subset is selected using limits of acceptability. Where do you see the differences?
- P20 L14-16: To me it is not clear from the context of the paragraph for which model (hydroecological model or hydrological model) the described data is used. Also, references to data sources are missing.
- P21 Figure 1: The last column of stage 1 uses three boxes to show the model output. I would suggest to reduce it to two boxes. This could either be i) one box for covariance (observed and simulated) and one box for HI (observed and simulated) or ii) one box for simulations (covariance and HI) and one box for observation (covariance and HI).
- P22 L7-9: Could you shortly state why exactly you selected an e function and not a linear function? How did you decide that you want 20 parameter sets and not e.g. 30?
- P22 Fig. 2: The figure could be complemented by a legend to make it easier to understand. I would also suggest to use the grey area to highlight the plausible parameter space and not the parameter sets not used for the final simulations.
- P24 L6-8: The range of parameter values is used as an indicator for consistency. However, the possible/plausible values of the model parameters vary in order of magnitudes. E.g. x4 that is an indicator for the event length (days) has different plausible values than x1 that is an

indicator for the storage (mm). This leads to your result that those parameters having values in the order of 100 have a much lower consistency than those having values in the order of 1-10. I would either normalize the parameter values or not use the range as in indicator for consistency.

- P 27 L9: NSE is a measure where the hydrological model simulations are evaluated against the simple model of predicting mean discharge for every day.
- Section 3.2.5: In this section you compare the efficiencies across your study catchments and you don't present any new results. So I would move the whole section into the discussion chapter.
- Section 4.1: This section is a summary of your results chapter. In the discussion, to me it would be more interesting to read about why you have certain results. You actually do that in section 4.3.1. I would therefore merge section 4.1. and 4.3.1.
- P31 L1: I don't think that your statement "… practical limit to the number of ER HIs" is correct. Because if one e.g. uses an optimization algorithm with a multi-criteria objective function, then each criteria can have a certain weight (equal weights or as in your case weights according to importance). So I don't think that the number of criteria is limited from a practical perspective

Technical details

- P18 L23: What does WFD mean?
- P18 L23-25: I think the sentence "Under the modified covariance…." doesn't fit into the context of this paragraph and I would remove it.
- P20 L1: I would rename the chapter from "data" to e.g. "Selection of ecologically relevant hydrological indicators", because most information in this chapter is on the selection process of the ER HI.
- P20 L1-10: You provide a description of the hydroecological modelling approach in the appendix. I would therefore shorten this paragraph to the minimum information needed to understand how you selected the ER HIs. Sentences such as "measure of the statistical weight…" or "Consequently, more conclusive statements may be made…" could be removed. However, I would complement the section with the information on how many ER HI were selected for each catchment, because this information is relevant for stage 3 in your modelling approach.
- P23 Table 3: I would extend the description of NSE: "…A measure of goodness of fit to the 1:1 line normalized by the variance". The detail on the normalization is important and you refer to in the discussion.
- P24 Table: Table 2 should be relabelled to table 4. I also suggest to add parameter x6 for completeness.
- Figures 3, 4, and 5: Please explain what the subplots titles, such as D, F, T, and R mean.
- I made the general observation that many Figures are not readable in black and white. Can you think of a way to change that?