Reply to reviewer 2

Antoine Pelletier, on-behalf of all co-authors

October 26th 2021

We would like to thank Reviewer 2 for their constructive comments on the manuscript. We provide an answer to each comment hereafter.

Specific comments

Comment: *I understand the original version of the GR6J model does not simulate groundwater levels, i.e. a state which might be comparable with observed groundwater levels? In order to make use of groundwater level information, the authors added a new module to the existing model, i.e. two more parameters to represent the relationship between one of the model stores and the normalized groundwater level time series. I am not sure why this form in equation (2) was chosen? What is the physical meaning behind? Of course, the authors performed a correlation analysis between groundwater levels and the states of the different stores/reservoirs, and the state of the exponential store correlated the best with measured groundwater levels – still I very much miss the physical reasoning behind this choice and generally, the methodology and model itself.*

Reply: The purpose of the work presented in the paper is to improve the ability of the GR6J model to simulate low flows, by adding a new source of information: groundwater level data. Explicitly modelling an aquifer-river relationship is a complex exercise, but we wanted to see whether the mere point groundwater level signal could be a useful source of information for hydrological modelling. The original version of the GR6J model does not simulate groundwater levels, since it is a conceptual model that was not initially foreseen to reproduce all physical hydrological processes taking place in the catchment. Therefore, none of the state of the existing model structure was designed to be comparable to observed groundwater level. However, we performed an empirical correlation analysis that showed that the exponential store is the most correlated to observed groundwater level; thus, we decided to use it to design our additional module to simulate groundwater level. Clearer explanations will be added in the 1.4 section of the revised paper. The choice of equation (2) was not led by physics, but, as we already replied to reviewer #1, we tried to use several formulations of polynomial relationships to transform the exponential store level into the normalised groundwater level, with degrees up to 3. It appeared that using a relationship with a degree 2 or more was not useful to improve performance; therefore, we decided to use an affine function. This information will be added in the revised version of the paper.

Comment: *On a very similar note, I was also wondering whether a generally good correlation between observed groundwater levels and the storages might have been caused by the choice/selection of catchments, i.e. those catchments were chosen where geology plays a dominant role in streamflow dynamics/runoff generation? – it might be good to see how well streamflow and the states of different stores are correlated with other fluxes and states, for instance precipitation, soil moisture, etc.*

Reply: Our selection of catchments was, according to the criteria exposed in lines 231–215, the widest possible in mainland France using available streamflow and groundwater level data. We agree that there may be a selection bias in our study, mainly because aquifers that are regarded as important for surface water resources — such as the chalk aquifer in Picardy — are the ones that have been monitored for a long time in a large number of measurement points. However, we think that our

catchment dataset is diverse enough to draw general conclusions, at least in climatic and geological conditions similar to Mainland France. This point will be exposed in the discussion section. About the other stores of the model structure, Aubert, Loumagne and Oudin (2003) showed that the production store is correlated to soil humidity and it could be used for soil moisture data assimilation.

Comment: *Generally, it would be good to add more physical reasoning to the manuscript, both in the methods and results and discussion sections. At the end of the results and discussion section, the catchments are categorised into 6 groups based on geological context, but the differences presented e.g. on Figure 17 are not explained in detail. Why can be streamflow/groundwater levels better simulated on one group than an other? What happens in reality in the catchments?*

Reply: The group analysis presented in figure 17 helps us produce the recommendations formulated in section 4.6 for model calibration in a practical context. However, it is impossible to know what happens in reality in the catchments, since they are not experimental instrumented catchments in which hydrological processes are monitored by in-field instruments. Due to the very limited amount of data we have, we did not want to make uncertain assumptions about hydrological processes taking place in the catchments.

Comment: *It is not mentioned whether the chosen piezometers are located on hillslopes or valley bottoms of the selected catchments – the groundwater level dynamics may be very different.*

Reply: It depends on the availability of groundwater level data. Since most studied catchments are lowland ones, most piezometers are located on high plains between valleys. While analysing the results of the study, we tried to make a distinction between piezometers according to their location but we did not find a satisfying way to distinguish between valley bottoms, hillslopes and high plains.

Comment: *I am not sure if strong conclusions such as general recommendations for model calibration processes could be drawn based on the findings of this study? The choice of the model structure, and also the catchments (e.g. eliminating snow melt dominated, etc. catchments) may very much influence the results.*

Reply: The recommendations formulated in section 4.6 are intended for the GR6J model in hydro-climatic and hydrogeological contexts that are similar to mainland France. They might be extended to other conceptual models with similar structures. This remark will be added to the revised version of the paper.

Comment: *Considering Lines 170-181, the large variety of hydrological and hydrogeological configurations – I am not sure why the same model structure is used for each catchment or why it is expected that using additional data but the same model structure, streamflow simulations may improve?*

Reply: The scope of the study is to build a general-purpose low-flow simulation model, usable in a large variety of hydrological and hydrogeological configurations. That is why the structure we propose is tested on a large sample of catchments.

Comment: *The description of the methods could be probably slightly improved, in a way that the results could be reproduced, e.g. adding more details on selection of catchments, filtering low quality piezometer data, assessing groundwater level dynamics, calculation of potential evapotranspiration (please see detailed comments below).*

Reply: These specific points will be explained in the revised version of the paper; they will be addressed below in the replies to technical corrections.

Comment: *Line 228: the authors argue that catchments with snow melt processes were eliminated from the study, therefore, no solid precipitation was simulated by the model. However, for certain catchments the maximum altitude is above 1000 m. How were these catchments handled? How was snow melt handled in these cases?*

Reply: Solid precipitation data are available in the SAFRAN reanalysis that provided the input climatic data for the simulations. As explained in line 206 of the manuscript, catchments with less than 10 % of solid precipitation were discarded. Several gauges located in the Alsace plain have part of their catchments in the Vosges mountains, which makes them reach maximum altitudes above 1,000 m, but the average altitude remains low enough not to overtake 10 % of solid precipitation. This explanation will be added in the revised version of the paper.

Comment: *Figure 4: It might be good to explain the model structure or at least the names of the variables in the main text if these variables are presented on a figure in the main text instead of the appendix. For a reader it might be confusing jumping between the main text and appendix in order to understand the methods.*

Reply: We did not want to add too much information about the model structure into the main text, in order not to confuse the reader. We agree that some essential information is missing; it will be added in the revised version of the paper.

Comment: *In case the authors decide to combine the Results and Discussion sections – it would be good to add further references to this combined section – how does this study fit into existing literature?*

Reply: As we already replied to reviewer #1, we will add more comparisons with previous studies in the discussion section.

Technical corrections

Technical corrections will be addressed in the revised version of the manuscript. Some of them need answers; they are listed below.

Comment: *Line 8: Abstract: it might be good to add a few numbers, what does "satisfying performance" mean?*

Reply: We agree that *satisfying performance* is a vague expression but we cannot add, for instance, an explicit value of median ZError in the abstract, since we would need to define ZError for that.

Comment: *Line 12: Subtitle – "low flow modelling" – the paper evaluates modelling of entire streamflow time series, including peaks, not only low flows – why are only low flows introduced here?*

Reply: The scope of our study is low-flow modelling, that is why the streamflow simulation performance are evaluated criteria focusing on low-flows, such as $NSE(\sqrt{Q})$ and $NSE(\sqrt[3]{Q})$.

Comment: *Lines 15: "complex water cycle underground processes" – this expression could be maybe revised?*

Reply: As we already replied to reviewer #1, the sentence will be rephrased as "The hydrological processes taking place underground, whose complexity is not straightforward to describe, are often aggregated in surface hydrology models by a simple reservoir, which fills during each rainfall event and slowly empties during rainless periods".

Comment: *Line 85: computation costs might not be an issue in certain cases considering advances in computer sciences?*

Reply: We agree that the problem of computation costs have been lowered by advances in computer science, that is why they are *"no longer mentioned in recent studies"*. But it remains an issue for some operational users that do not have access to state-of-the-art computation facilities; thus, we think it needs to be mentioned.

Comment: *Line 98: do the authors mean surface runoff/rainfall-runoff modelling studies, i.e. not most studies?*

Reply: The sentence will be clarified in the revised version of the paper.

Comment: *Line 121: I am not sure what is meant by "but a visual evaluation of calibration is necessary"?*

Reply: The calibration of the model is not performed automatically, but the modeller must adapt the calibration "by hand" in the case of this study to fit both objectives.

Comment: *Line 129: the meaning of this sentence is not clear: 1) it might be good to split it into shorter sentences; 2) what is meant by "in a catchment which few streamflow measurements" – do the authors mean "with few streamflow measurements"?*

Reply: The word "which" is a mistake of our own; it will be replaced by "with" in the revised version of the paper. Thereby, the sentence will be clearer.

Comment: *Line 154: I am not sure, perhaps using piezometric information, streamflow simulations usually deteriorated?*

Reply: In the revised version of the paper, the expression "had little influence" will be replaced by "did not significantly impact".

Comment: *Line 200: RNESP is heritage national network for groundwater monitoring? If yes, it might be good to remove "-" twice or replace the second with a comma.*

Reply: In the revised version of the paper, the punctuation of this sentence will be made clearer.

Comment: *Line 203: please explain on which basis these catchments were selected?*

Reply: The catchments were selected using the criteria exposed in section 2.3, on the basis of data availability. This sentence will be rephrased in the revised version of the paper.

Comment: *Line 203: what is meant by "relevant"? Please explain.*

Reply: As explained in lines 207–209, a piezometer is regarded as relevant if its screen is located within an aquifer emerging inside the catchments boundaries. This sentence will be rephrased in the revised version of the paper.

Comment: *Figure 2: Hydrological maps? These seem to be geological maps or hydrogeological? Seno-Turonian chalk and Late Cretaceous multi-layer limestone seem to have very similar colour – is it possible to change this? It might be good to add a small map of France to this figure and indicate the locations of the sample catchments within France also on this figure.*

Reply: The word "hydrological" is a misprint of our own, it will be replaced by the word "hydrogeological" in the revised version of the paper. The locations of the two catchments are shown on figure 1, which is on the previous page; we preferred to show them on a separate figure instead of adding more information to figure 2.

Comment: *Line 211: please explain what "too low quality" means? Is there any objective/automatic way which was used to perform this step – a way readers can use to reproduce the results?*

Reply: We did not manage to perform this step in an automatic way. We eliminated data that were clearly affected by noise. We also relied on the expertise of database maintainers who helped us select the time series (they are thanked in the acknowledgements section). A short explanation will be added in the manuscript.

Comment: *Line 208: how was the relative importance of hydrogeological formations assessed? Importance in/for what? Please explain in the manuscript.*

Reply: As we already replied to reviewer #1, the relative importance of each hydrogeological formation was assessed visually, using maps similar to the ones in figure 2, to eliminate formations whose outcropping or sub-outcropping areas represented less than 5 % of the catchment area. Very few piezometers have been eliminated this way; those which have been discarded for this reason were often located on the wrong side of underground watersheds, as identified by BDLISA, when these watersheds differ from the limits of the catchment. This will be clarified in the revised version of the paper.

Comment: *Line 220: how exactly were these dynamics compared? Please explain in manuscript*

Reply: These dynamics were compared visually, on the whole time series and using Pearson correlation. This will be added in the revised version of the paper.

Comment: *Table 1: does surface mean surface catchment area?*

Reply: Yes, this will be clarified in the revised version of the paper.

Comment: *Table 2: PET stands for potential evapotranspiration?*

Reply: Yes, this will be clarified in the revised version of the paper.

Comment: *Table 2: please add to the methodology section how potential evapotranspiration was calculated.*

Reply: As mentioned in line 194, potential evaporation was computed using the formula by Oudin et al. (2005). This will be added to the revised version of the paper.

Comment: *Line 246: this sentence seems to be incomplete.*

Reply: The sentence will be rephrased in the revised version of the paper.

Comment: *Line 250: I understand solid precipitation was not simulated in the manuscript – if this is the case, this sentence might be confusing in the Methodology section.*

Reply: This paragraph will be removed from the revised version of the paper.

Comments: *section 3.1.2: in the Methodology section please describe only that objective function/s which was/were used in this manuscript.*
Line 261: but which one was used in this manuscript? Please only describe/mention that one which was used.

Reply: This section intends to describe the usual calibration strategy of the GR6J model. We think that it is important for the reader to understand how original is the calibration scheme implemented in this study, described in section 3.4. A clearer explanation will be added in the revised version of the paper.

Comment: *Line 267: Please add here which are these three, perhaps describe a bit these reservoirs (e.g. how they work, physical meaning, etc.)*

Reply: The equations that rule the stores are described in appendix C, but a short explanation will be added to this section.

Comment: *Equation 1: the term anomaly might be confusing – according to equation 1 – groundwater level time series were simply normalized?*

Reply: We agree that the word "anomaly" may be confusing. We will remove it from the revised version of the paper.

Comment: *Line 316: effect on what? Please explain*

Reply: The word "performance" will be added to the revised version of the paper.

Comment: *Section 3.5: please add exactly which years (how many years) were involved in P1 and P2?*

Reply: The exact years depend on the catchment/piezometer pair considered, since it is a function of the availability of data. With the selection criterion exposed in line 215 of 10 years of continuous contemporaneity between streamflow and groundwater level data, the theoretical minimum duration of P_1 and P_2 is 5 years. In the sample, durations of periods are from 5.6 years to 28.5 years, with a median at 16.4 years and a mean at 17.0 years. This information will be added to the revised version of the manuscript.

Comment: *Figure 6: top panel: are there values (below 0) which were cut off? If yes, maybe it is better to indicate e.g. in figure caption.*

Reply: Yes, these values were cut off for readability. This will be added to the revised version of the paper.

Comment: *Starting with line 350: this belongs to the Methods section.*

Reply: This will be moved to the Methods section in the revised version of the paper.

Comment: *Line 350: please explain here what the authors mean by "differences between evaluation criteria distributions"?*

Reply: In figures 6 and 7, the distributions of several criteria are shown. These criteria are used to evaluate the model performance in several scopes: the ensemble of the hydrograph, low flows and groundwater level. Distributions are shown as boxplots in these figures, which allows the visual evaluation of the differences between distributions. For a given criterion, if the difference is globally positive — respectively negative — with respect to the reference, it means that the model performance is increased — respectively decreased.

Comment: *Line 393: please explain what is meant by transferability?*

Reply: Transferability is the possibility to use parameters for a simulation period although they were calibrated on another one. A short explanation will be added to the revised version of the paper.

Comment: *Line 401: please explain what is meant by "direct spatial pattern"?*

Reply: The word "direct" will be replaced by "clear" in the revised version of the paper.

Comment: *Line 403: "High scores" – please rephrase.*

Reply: In the revised version of the paper, the word "scores" will be replaced by "values of performance criteria".

Comment: *Lines 406-408: please add why?*
Lines 409-416: please add more physical reasoning; and explanations why these results were found.

Reply: We searched the hydrogeological literature for explanations of the result patterns and we found no useful explanatory variable, would it be aquifer porosity, transmissivity, thickness or any other aquifer characteristics. Therefore, we cannot explain why these results were found in terms of physics, all the more so as our model, as explained above, does represent hydrological processes in an aggregated — *conceptual* — way.

Comment: *Page 23: this text belongs to the Methods section*

Reply: The description of the sub-group analysis will be moved to the Methods section.

Comment: *Section 4.6: please add that these findings are for the specific case of using this model*

Reply: This will be added in the revised version of the paper.

Comment: *Lines 461-474: this belongs to the Discussion section.*

Reply: As already explained in the reply to reviewer #1, the Discussion and Conclusion sections will be re-organised in the revised version of the paper.

Comment: *Line 448: what does probably mean? Did the authors reduce it or not?*

Reply: We used the word "probably" because we did not explicitly assessed equifinality, since it would require a whole ad hoc study. This sentence will be rephrased in the revised version of the paper.

Comment: *Line 455: "groundwater resource management anticipation" – I am not sure what this means, is it possible to rephrase?*

Reply: This expression will be rephrased as "an anticipative management of groundwater resources".

Comment: *Line 475: what does particular conditions mean? Please add some information on this.*

Reply: The particular conditions are detailed on the Banque Hydro website. They are being changed since the Banque Hydro is moving to its third version. Any further explanation included in the paper would certainly become obsolete in a few months. Therefore, we do not think that it is suitable to add more information, since it can be found on the website.

Comment: *Line 484: does this mean that actual evapotranspiration is not calculated in the model?*

Reply: Actual evaporation is the variable E_n, which is computed at every time step in the model.

References

Aubert, David, Cécile Loumagne and Ludovic Oudin (Sept. 2003). 'Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model'. In: *Journal of Hydrology* 280.1-4, pp. 145–161. DOI: 10.1016/s0022-1694(03)00229-4.

Oudin, Ludovic et al. (Mar. 2005). 'Which potential evapotranspiration input for a lumped rainfall–runoff model?' In: *Journal of Hydrology* 303.1-4, pp. 290–306. DOI: 10.1016/j.jhydrol.2004.08.026.