Reply to reviewer 1

Antoine PELLETIER, on-behalf of all co-authors

October 8th 2021

We would like to thank Reviewer 1 for their constructive comments on the manuscript. We provide an answer to each comment hereafter.

Specific comments

Comment: *L55-56: Conceptual and physically-based models are mentioned representing less and more physical processes in a model. Shouldn't empirical hydrological models be mentioned here as well, since these present the 'lower' end of the 'physics spectrum'?*

Reply: We agree with the fact that empirical models should be mentioned here beside conceptual and physically-based models. This was an oversight of our part. It will be added in the revised version of the paper.

Comment: *L118-120: This sentence is not very clear and needs to be reformulated. Furthermore, it is doubtful whether the apparent content generally is valid. At least when compared to simpler, conceptual hydrological models, it will be more difficult to match simulated and observed time series of physical variables in physically based models, since more parameters are fixed based on observations – hence reducing the degree of freedom in the calibration – and calibration processes are cumbersome and time consuming.*

Reply: The purpose of these sentences was to introduce a short review of the literature in which observed data – in situ or remote-sensed measurements – were used to calibrate physically-based models. We wanted to underline that it is, at least when designing the calibration process, easier to use these data when the physical variable is explicitly simulated by the model, than when it is not and you have to build a function between a variable within the model and the observational data. Since physically-based model generally have more state variables and thus, they simulate more physical variables, designing the calibration process is generally easier. However, computation costs are often much heavier than for conceptual models. We agree that this section is not clear enough and it will be rephrased in the revised version of the paper.

Comment: *L192: Which climatic data are meant here and which climatic data were used to calculate the potential evaporation?*

Reply: Climatic data taken from the SAFRAN reanalysis are daily precipitation and average temperature. The daily rate of solid precipitation was also used to eliminate catchments that are affected by snow. This information will be added in the revised version of the paper.

Comment: *L208-209: How was the relative importance of each hydrogeological formation assessed?*

Reply: 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 precised in the revised version of the paper.

Comment: *Table 2: The variability of the mean annual potential evaporation is quite low (range of 600-792 mm), where I would expect a much larger range for France given the diversity in climatic and geographical conditions. Could the authors explain this a bit more?*

Reply: This is not uncommon in France, especially when no catchment with a Mediterranean climate is included in the catchment sample. Catchments located in the southeast regions of Mainland France can reach an annual potential evaporation above 800mm (824mm for the Huveaune river in Aubagne; 842mm for the Mosson river in Saint-Jean-de-Védas; 880mm for the Réart river in Villeneuve-de-la-Raho), but none of them is part of the study dataset. Values under 600mm are reached in mountainous catchments (538mm for the Doubs river in l'Abergement-Sainte-Marie; 319mm for the Arve river in Chamonix-Mont-Blanc) but they were discarded from this study to avoid taking into account solid precipitations (Brigode et al. 2021).

Comment: *L268-269: How has GR6J been calibrated; which optimization method has been used? And which data and time period(s) were used in this stage of the study?*

Reply: For each catchment, GR6J was calibrated on the whole period of climatic data (1958-2018) using a calibration algorithm made on purpose by Michel (1991). This algorithm gives similar results as the differential evolution one used for the main part of the study, but its computation cost is much lower, that is why it was used for the ordinary calibration on streamflow. This explanation will be added in the revised version of the manuscript.

Comment: *L291: Why has equation (2) been used to transform the exponential store level to the normalized piezometric level? Which alternative relations have been investigated and which criterion has been used to select this particular equation?*

Reply: 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: *L300-304: The conversion of equation (4) to equation (5) does not seem to be correct. The observed and simulated piezometric anomalies can each be expressed according to equation (1). Combining equation (1) and (4) does not result in equation (5). For instance, one would expect to see the observed and simulated standard deviations and the average simulated value in equation (5). This does not necessarily disqualify ZError as expressed by equation (5) and used in the model calibration and validation, but the derivation of equation (5) should be reconsidered.*

Reply: We understand that the explanation in this part is not clear enough. In fact, the model structure that we propose does not simulate absolute groundwater level, but only its normalised version. To compute simulated groundwater level, it is necessary to multiply the simulated anomaly by the standard deviation of observed data and to add the mean of the latter, which is expressed in equation (3). By combining equation (1), (3) and (4), we get:

$$ZError = 1 - \sum_t \left(\frac{z_{sim} - \bar{z}}{\sigma_z} - \frac{z_{obs} - \bar{z}}{\sigma_z} \right)$$

Which gives:

$$ZError = 1 - \frac{\sum_t (z_{sim} - z_{obs})^2}{\sigma_z^2}$$

By definition of standard deviation, we have:

$$\sigma_z^2 = \sum_t (z_{obs} - \bar{z})^2$$

And we get equation (5). This explanation will be added in the revised version of the manuscript as an appendix.

Comment: *Figure 7: An alpha value of 0 was discarded since no groundwater level simulation is 'performed' in that case. Do you mean that the model could not generate an exponential store level? The alpha value only is a weight is the composite objective function and hence should not influence the model simulations, isn't it? In addition, I think it will be interesting and relevant to compare the groundwater level simulations with the original calibration as well. Now validation results for the original and composite calibration are only compared for streamflow (e.g. Figure 16), but a similar comparison for groundwater levels seems to be relevant as well. What is the improvement in groundwater level simulation when taking piezometry into account in the calibration compared to the traditional approach where only streamflow data are used?*

Reply: The GR6J model only uses six parameters are used to simulate streamflow in the model structure; the seventh and the eighth parameters are added to simulate groundwater level anomaly. Therefore, when calibrating with the original calibration strategy – i.e. $\alpha = 0$ – the sensibility of the calibration criterion to parameters X_7 and X_8 is zero and they are randomly determined by the stochastic optimisation algorithm. Thus, we can say that in this case, no relevant groundwater level anomaly is simulated, except a random affine transformation of the exponential store level which cannot be relevantly compared to observed data. This explanation will be added in the revised version of the paper.

Comment: *L430-474: Section 4.6 (Synthesis) and section 5 (Conclusions) both contain conclusions and partly discussion. Try to strictly separate discussion of limitations, comparison with other studies and generalization issues (Discussion) from the main findings linking to the objective of this study (Conclusions). In addition, the discussion section can include some more comparisons with previous studies (e.g. studies mentioned in the introduction), where the value of data in addition to streamflow data for calibration and validation of hydrological models has been assessed.*

Reply: Sections 4.6 and 5 will be re-organised in the revised version of the paper to separate discussions from conclusive remarks. We think that synthetic guidelines presented in section 4.6 are likely to be a key message of the paper for future users of the model and therefore, they will be highlighted in the conclusion. We will also add more comparisons with previous studies.

1 Technical corrections

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

Comment: *L4: 'lumped rainfall-runoff models'; this term indicates the spatial aggregation scale of the model, but does not give information on the extent to which physics are incorporated in these models (i.e. empirical, conceptual and/ or physics-based models).*

Reply: The word "conceptual" will be added in the revised abstract.

Comment: *L15: 'complex water cycle underground processes'; what do the authors mean with this term?*

Reply: 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: *L20-21: 'hydrogeological' (line 20) and 'geological' (line 21); has this distinction been made on purpose or should a consistent term be used?*

Reply: "Hydrogeological" will be used.

Comment: *L86: 'used a groundwater reservoir'?*

Reply: The "ground reservoir" is the name of the reservoir inside the structure of Gardénia model. In the revised version of the paper, it will be put in italics.

Comment: *Table 2: What is the definition of catchment yield?*

Reply: Catchment yield is defined as the quotient of mean annual streamflow divided by mean annual rainfall, both of them expressed as water depths. This definition will be added in the revised version of the paper.

Comment: *L287: What is the meaning of z?*

Reply: z is the absolute groundwater level, expressed in meters above sea level. This will be added in the revised version of the paper.

Comment: *L394: 'as the result of fluxes between topographic catchment'; what is meant here?*

Reply: The exchange function evoked here takes into account the fact that in most cases, the topographic catchment of a gauge station – i.e. the one computed using a digital elevation model and its drainage network – does not exactly correspond to its real catchment – i.e. the set of points that actually contribute to streamflow. Therefore, topographic catchments lose or earn water from their neighbours and the exchange function represent theses fluxes. A full discussion about this point is available in Le Moine (2008). A short explanation will be added in the revised version of the paper.

Comment: *L469-471: This explanation is hard to follow, please try to rephrase.*

Reply: This convoluted explanation will be rephrased in the revised version of the paper.

References

Brigode, Pierre et al. (2021). *Summary sheets of watershed-scale hydroclimatic observed data for France.* INRAE. URL: `https://doi.org/10.15454/UV01P1`.

Le Moine, Nicolas (Nov. 2008). "Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des modéles pluie-débit ?" Thèse de doctorat. Agro ParisTech.

Michel, Claude (Sept. 1991). *Hydrologie appliquée aux petits bassins versants ruraux.* Cemagref. Antony, France.