

Reply to anonymous referee 6

We would like to thank the anonymous referee 6 for his comments on the manuscript and we are glad he finds this work presents an interest for the hydrogeological community, and that he deems the paper acceptable for publication after a minor review. We propose an answer for each specific comment hereafter.

Comments:

1. *The manuscript acknowledges that a proper evaluation of the contribution from the hydrogeological component would require a multi-model approach, which is not available for most of the AquiFR domain. However, it subsequently concludes that the contribution of the hydrogeological components to the total uncertainty remains negligible compared with that from the climate components. I may suggest authors stating more explicitly that hydrogeological-model uncertainty is not explicitly quantified in this study and moderating the corresponding conclusion. In addition, the subsequent comparison of the CDFs mainly evaluates the influence of the climate projections on the historical reference distributions used to construct the SPLI, rather than directly evaluating uncertainty arising from the hydrogeological models. The scope of this analysis should therefore be clarified accordingly.*

- Reply: the referee is right about the fact that a proper evaluation of the contribution from the hydrogeological component would require a multi-model approach. This fact is stated at the lines 427-428, along with the fact that such procedure could not been carried out because AquiFR is not based on a multi-model approach. Indeed, the authors might not have stated this point clear enough. We propose to rephrase the sentence at lines 431-433:

“Thus, we can assume that models with spatial overlap in the AquiFR region may yield comparable results, and consequently that the contribution of hydrogeological components in the modelling chain to overall uncertainty is relatively minor compared to that of climate components. However, as stated above, a thorough assessment of the hydrogeological contributions to total uncertainty is necessary to substantiate this claim.”

Furthermore, the referee is right in saying that the comparison of the CDFs over the historical period does not properly assess the uncertainty contribution related to hydrogeological components. To better emphasise this fact, we propose the rephrase the lines 433-436:

“While this kind of assessment remains unfeasible, we can nevertheless evaluate the influence of climate projections on historical reference distributions of groundwater levels. Indeed, part of the uncertainty is linked to the fact the climate projection cannot be fully unbiased over the historical period, which might affect simulated groundwater levels. This is rather explicit when computing SPLI over the historical period, since the projections are not compared to present day observations or reference simulation, but to present day climate simulation, which are all different.”

2. *In Section 2.3.2, the Mann–Kendall test is currently cited as Mann and Whitney (1947), which might be a wrong citation. Please check it. More importantly, detailed information is needed to verify which statistical test was actually implemented in the analysis and clearly specify the input time series and temporal resolution used for both tests, whether the tests were applied to the*

original time series or to the 5-year running median. If necessary, the significance results reported in Appendix C should be reassessed accordingly.

- Reply: as requested, we changed the reference corresponding to the Mann-Kendall test at lines 269-270 as follows:

“The Mann-Kendall test (Mann, 1945; Kendall, 1948) on the future SPLI values is used to assess whether the associated trend of future SPLI values significantly evolves over time.”

Regarding the statistical tests, we propose to add the requested information when describing the used tests:

- Kolmogorov-Smirnov test at lines 267-268:

“The test is performed using the Scipy Python package, and the null hypothesis is that distributions are identical;”

- Mann-Kendall test at lines 269-270:

“The test is performed using the pyMannKendall Python package, the null hypothesis is that there is no monotonic trend in the future SPLI values.”

Regarding the inputs on which the tests were performed, as mentioned at the lines 257, the tests are performed on the temporal evolution of the spatial proportion of the Aquifer surface area affected each category of events. In practice, the SPLI values from all models are regrouped and the proportion of surface affected by a specific return period is defined as the sum of each pixel's surface dedicated to this return period, and divided by the total surface area of the Aquifer domain (as mentioned at the lines 258-260). Therefore, we propose to not add additional information on this point. However, we agree that the fact that the tests are performed on the original time series or on the 5-year temporal rolling median remains unclear. We propose to add the following information by completing the sentence at the line 273:

“For each storyline, the tests are performed on the 5-year temporal rolling median from the corresponding GCM/RCM couple. When considering all the CPs, in order to provide more integrated analysis for each RCP, the tests are performed on the median from the 5-year temporal rolling medians over the available GCM/RCM couples. For each statistical test, the significance threshold is set at a p-value ≤ 0.05 .”

3. The Introduction states that Aquifer “is not based on a multi-model approach” because each hydrogeological modelling software package is generally applied to a specific catchment and very few simulated areas overlap spatially. In contrast, Section 2.2 describes Aquifer as “a hydro-meteorological modelling platform based on a multi-model approach.” It is recommended to clearly distinguish between a platform that integrates several regional hydrogeological model configurations and a hydrogeological multi-model ensemble that can be used to assess structural model uncertainty, and then use the terminology consistently throughout the manuscript.

- Reply: the referee is right, this point is unclear. The Aquifer platform is not based on a multi-model approach, it is more like an ensemble of regionally configured hydrogeological models. We propose:
 - To rephrase the paragraph lines 159-161 of the latest reviewed version accordingly:

“Aquifer is a hydro-meteorological modelling platform regrouping existing models, one model being defined as a pair of catchments and the hydrogeological modelling software used, to simulate French rivers and groundwater systems (Habets et al., 2021).”

- To delete the expression “based on a multi-model approach” at the lines 74-75;

4. A systematic cross-check of the quantitative information in the main text and Supplementary Material is recommended. For example, the model-area range reported in Section 2.2.3 is inconsistent with the values in Table A2; the spatial resolutions of the GCMs and RCMs are expressed in km² rather than km; and the third future horizon in Tables C1–C3 appears to contain an incorrect time-period label. In addition, the origin of the “6.3%” value reported for the green storyline in Section 3.2 should be checked against Table C1, and the statement that extreme wet events increase “by a factor of 7” in Section 4.3 does not appear to follow directly from the values reported in Table C3. I therefore recommend checking the relevant values, units, time periods, and calculations throughout the manuscript and Supplementary Material.

- Reply: as requested, a systematic cross-check of the quantitative information between the main text and the supplementary material was carried out:
 - Rephrasing the paragraph at lines 218-220:

“Two other models are located further out: MARTHE Alsace covering an area of around 1,100 km² in East of France, and MARTHE Tarn-et-Garonne, of a similar size, which is the only model located in the southern part of mainland France. For the remaining models, the surface areas of the aquifer systems within the models vary between around 7,000 and 66,000 km².”
 - Replacing the “km²” by “km” when describing the spatial resolution of the GCMs and RCMs in section 2.1;
 - Checking the values between the main manuscript and the supplementary material;

5. Section 2.3.1 explains that monthly CDFs are fitted using a non-parametric Kernel Density Estimation with the scikit-learn package. However, the main KDE settings, particularly the kernel, bandwidth, and bandwidth-selection procedure, are not specified. These details should be briefly provided to improve reproducibility. In addition, it is recommended to use precise terminology in the Results and Conclusions instead of repeatedly referring simply to “10-year wet events” or “10-year dry events”, as was done in the “Methods” section, where the SPLI values were correctly related to the return period of the equivalent event in the historical reference period.

- Reply: as requested, the information regarding the KDE has been added at the line 244:

“The KDE is estimated using the Epanechnikov kernel (i.e., with a parabolic form) and the bandwidth-selection procedure controlling the smoothing follows the Silverman's rule of Thumb, to get an optimized estimator of bandwidth.”

Regarding the use of precise terminology to describe 10-year events, we propose to replace all the expressions “decadal events” and the “10-year events” by “extreme events” in the Results and Conclusions section as extreme events are defined as at least 10-year events in this study, as established at the lines 278-279.

6. Please clarify whether Figure 4 and Appendix C represent mutually exclusive SPLI categories or cumulative exceedance thresholds. Table 2 defines separate categories such as “between 2.5-year wet and 5-year wet,” “between 5-year wet and 10-year wet,” and “>10-year wet.” In contrast, Tables C1–C3 use expressions such as “≥2.5-year wet,” “≥5-year wet,” and “≥10-year wet,” which appear to represent cumulative exceedance thresholds. The authors should ensure that the definitions and terminology are consistent among Table 2, the Methods, the Results, figure captions, and Appendix C.

- Reply: The referee is right about the inconsistency between the labels of events in Table 2 and Tables C1–C3 in supplementary material. The SPLI categories are mutually exclusive as described in Table 2. The labels in Tables C1–C3 have been corrected accordingly. The row titles on Figure 4 have been corrected accordingly. Also, we propose to complete the line 250 to highlight this point:

“The range of SPLI values associated with return periods (T) is divided into seven mutually exclusive categories (Table 2).”

7. The Discussion appropriately acknowledges that the limited representation of southern France in the Aquifer platform prevents the existence of a France-wide North-South divide from being established with certainty. However, the Abstract still refers more broadly to groundwater changes “over France,” “South of France,” and a “discernible North-South distinction.” It is recommended to make the Abstract, Results, and Conclusions consistent with the more cautious interpretation in the Discussion, for example by referring to the Aquifer domain or to the French aquifer systems represented by Aquifer.

- Reply: as requested we propose the following changes:
 - Correcting the sentence at lines 2-3:

“As part of the EXPLORE2 project, the aim of this study is to estimate changes in groundwater levels on French aquifers during the 21st century.”

- Correcting the sentence at lines 14-15:

“This is accompanied by an increase in the 10-year low groundwater levels mostly observed in the southern part of the French aquifer systems represented by Aquifer, which is shared by the four storylines.”

- Correcting the sentence at lines 236-237:

“Therefore, the SPLI is widely used by the French Geological Survey (BRGM) over France, as well as for seasonal forecasts with Aquifer for the French aquifer systems represented by Aquifer.”

- Correcting the sentence at lines 507-509:

“Indeed, the conclusions drawn for the northern part of the Aquifer domain are not consistent with previous studies which, as mentioned in section 1, suggested a global decrease of both recharge and groundwater level over France [...].”

8. The Introduction currently states that the use of the selected projections as storylines to illustrate climate-change impacts on groundwater is an approach that, “to the best of our knowledge, has not

yet been carried out in the literature.” Unless this statement can be supported by a sufficiently comprehensive literature assessment, I suggest avoiding such an absolute claim.

➤ Reply: as requested, we corrected this statement at the lines 80-82:

“These projections are utilised as storylines to explore the range of climate change impacts on groundwater, an approach that, to our knowledge, has been rarely adopted in existing literature.”

References:

Kendall, M. G.: Rank Correlation Methods, Charles Griffin and Co., London (UK), 1948.

Mann, H. B.: Nonparametric Tests Against Trend, *Econometrica*, 13, 245–259,
<https://doi.org/10.2307/1907187>, 1945.