

We would like to thank the two reviewers who kindly provided us with their comments.

The initial objectives described in the introduction were to present “the methodology underlying the dataset, the evaluation of the hydrological models against daily streamflow and groundwater level observations, the assessment of the future streamflow and recharge projections, the data availability and the ways of accessing the data and understanding the results (mainly through visualisation tools)”. These objectives were too ambitious for one paper. It is clearly unsatisfactory and the content of the different sections is finally unbalanced. The article is already too long. Based on the recommendations of the two reviewers, we have chosen, in the next version, to reorganize the paper, to complete the sections on the model evaluation procedure and the description of the results, to add a literature review and to reduce the section devoted to the transfer of data and project results to stakeholders. The transfer to users will be the subject of a future, much more comprehensive article (including visual tools, the technical documentation and the interpretation of summary sheets).

We have paid more attention to the numbering of tables, figures and sections in the next version. This poor numbering made the manuscript difficult for reviewers to read. We apologise for any inconvenience caused and undertake to rewrite the sentences and avoid redundancies.

RC1: 'Comment on egusphere-2025-1788', Anonymous Referee #1, 17 Jun 2025

General comments

Sauquet et al. presented in their manuscript the results of French national project called Explore2 to provide hydrological climate projections in France for streamflow and groundwater changes until the end of this century. It is a successor project of Explore2070 which took place in the early 2010ies. The motivation to update projections were to extend projections to the end of the century with newer climate model data and to increase the locations for projections. Many institutions were involved over several years from science, administration and stakeholders in water management. The objective of the project also to provide a better accessibility and understanding of the data for stakeholders compared the predecessor project. The main objectives of the manuscript were to describe the ensemble of the simulations, to evaluate the ensemble in past climate and present main results of the climate projections.

They used a subset from the EURO-CORDEX CMIP5 climate chain ensemble, two bias correcting methods, nine surface hydrology models and four groundwater models. Models were evaluated in the past both using observed forcing and climate model data at large number of observation location, i.e. > 600 gauging stations for surface hydrology models > 200 piezometers for groundwater models over many years. Some of the models were calibrated using similar data as for evaluation. Not all models cover whole France. The authors present evaluation under multiple aspects and mention thresholds for acceptable performance.

The projections cover up to approx. 4000 simulation points for streamflow projections covering whole France and parts of France for groundwater recharge projections. Results for evaluation and projections focused on the surface hydrology models as results of the groundwater models are (or will be) presented elsewhere. However, there is a sufficient overview of the groundwater part presented in this manuscript.

The substantial content of the manuscript is, however, not matched with the quality presenting the data. Most importantly I miss a literature review on the current knowledge of hydroclimatic projections in France (and Europe). The question must be answered, to my opinion, how the presented results embed in and expand existing knowledge.

The second largest issue is the readability of the manuscript, which suffers under many aspects. Most problematic are the large number of the results, the imprecise language which requires the reader to interpret what could be meant and the manuscript structure. To my opinion the manuscript will largely improve with using a commonly used structure to separate data/methods and results/discussion. I will detail the arguments below. With this impression I recommend to major revisions for this manuscript.

Detailed comments

Literature review

As mentioned above the manuscript is not largely referring to previous work. This should also include regional studies. This aim was mentioned in the overview of the manuscript to be part of section 7, but I could not find a relevant discussion with previous literature. This literature review should also include model decisions (e.g. the impact of bias correcting methods ...). Section 7 rather adds new own results not presented before.

→ *A review of the literature on impact studies concerning France has been published in French (<https://doi.org/10.57745/XKMLMG>). This is a part of the technical documentation of the Explore2*

project. We have summarise the content of this report in the introduction and compare the results obtained here with the conclusions of the identified previous studies (see §7 L511-525).

Readability

First is the large number of results. The objective was to show only main results, and I think this can be more condensed. The large number of results in the figures and tables is supplemented with additional numbers in the text, which makes it difficult to read (e.g. on page 28). Generally, I advise that these numbers in the text could be presented in a table, however, given the need to condense results for this manuscript, I would vote for a more qualitative text description to put the chosen results in tables and figures into value.

→ We have reduced reduce the number of Tables (e.g., Table 1 in the previous version was redundant with Fig. 1 and was deleted). We have followed the reviewer's advice by removing the numbers from the text to make the article easier to read.

Second is the imprecise language. As it is a complex data set covering many dimensions it is essential to be precise. The reader may be able to follow what is meant but it requires a very thorough reading. A table and a figure, however, should be self-explanatory to a certain extent. This is often not given. As an example for many other parts is the presentation of Table 6. The unit is not provided, so one cannot say if the numbers are percentage changes or absolute changes. It can only be assessed from the context that the “interquartile range of the median change ...” in the caption description refers to an interquartile range between simulation points (space) and the median change to a set of simulations varying in climate models, bias corrections and hydrological models.

→ We have sought to clarify the nature of the statistics (overall, spatial, etc.).

Additionally, I could often not follow what the authors interpret from figures, either by wrong references to figures or that they were imprecise with mentioning patterns or regions. For example, in lines 482-483 “The most significant and positive changes with an agreement between projections can be found for QDJF in the Alps and Pyrenees (Fig. 8) for the three RCPs.” I could follow this statement with looking at Figure 7. Similarly, I could not follow the impressions of the elasticity values (lines 338-342) shown in Fig. 3: Are the values for temperature good, or for precipitation in the summer?

→ Please accept our apologies for any errors in the references to figures, sections and tables. We have made additional comments on elasticities (§ 5.2.1 L321-329).

Manuscript structure

The reading flow is often interrupted by switching between methods and results as the manuscript is more structured by content (evaluation, future projections ...) than separating data/methods and results/discussion. For example, the results on projected changes start with a description of a multi-model index of agreement (Eq. 2). Ideally the manuscript would keep the content structure but replicate this structure for the data/methods section and the results/discussion section.

→ We have reorganised the manuscript in accordance with these recommendations. The data used for the hydrological modelling will be presented in the first two sections (first: observations and meteorological reanalysis, and then: climate projections). The section on the co-construction process with stakeholders was shortened and incorporated into the conclusion. Methods and results are now presented in different sections.

Focus of the manuscript

In the abstract the authors mention prominently that they want to present the data availability and understanding the results through visualisation tools. This part is in the manuscript rather short. If the authors want to present their data availability, the understanding thereof and their visualisation tools as promised in the abstract I think this needs to be more elaborated in the manuscript.

For example, the fact sheet in the supplemented is quite complex and is not put in value in the manuscript text. I would have liked to see more details on the description of these sheets and how these sheets help to make informed choices pre-select hydrological models (mentioned in section 5).

The data availability is only presented with a link (section Data availability). There is no guidance in the manuscript or on the webpage how I can get netCDFs of a certain figure for example. The webpage is available in an English version but 99% of the text is still in French. Besides the language barrier (which can be solved with translation tools nowadays) the data is hidden in multiple links.

There is a link for each model for each RCP for each subregion, which leads to a new webpage with new links with filenames, which relate to different climate chains and bias correction methods. Similarly, the visualisation tool is presented just with a link but not put in value in the manuscript In a similar note: In the conclusion in line 595 the authors advertise their data set with the ability to step back from the results in case of disagreement and promoting an understanding of the differences between hydrological models. This is indeed possible with such a data set, but it is not presented in the manuscript. The authors show mainly median results between hydrological models (except of a minor subtopic with Figure 5). In Table 3 there are thresholds presented for model performance, but nothing mentioned about the consequences. It would be interesting if the authors would include a detailed description of their fact sheets and how one could potentially use them.

→ *Based on the two reviewers' comments, we have understood that we were overly ambitious in attempting to present both the Explore2 dataset and the main findings, as well as how stakeholders can use the information provided by the project (including model selection). The transfer of knowledge from the project will be the subject of a separate article. Thus, the formerly section entitled "Co-construction between scientists and stakeholders" has been shortened and moved to the conclusion section (L582-599).*

Minor points

Reply

- In the abstract there are no results summarized for the groundwater models
⇒ *The following finding has been added: "In addition to northern France, annual groundwater recharge is projected to increase slightly in the north-east and remain unchanged elsewhere by the end of the century, according to the RCP8.5 scenario."*
- Section overview is inconsistent with section headlines and content (lines 91-95). Similarly, reference is probably wrong in line 446 (Table 7 is not existent), and section references (line 426).
⇒ *Sorry again for the wrong numbering.*
- Line 103: What are the criteria selecting the subset of climate models? The cited study, which could answer the question is only available in French.
⇒ *The criteria used for selecting GCM/RCM are related to data availability, model diversity and simulation quality and are detailed in § 3 (L127-139).*

- Line 111: Why is the reference period (sometimes called with different names, try to use consistent naming for better readability) chosen to be 1976 to 2005 opposite to the latest AR5 and AR6 reports? This could be better motivated.
 - ⇒ *The term “benchmark period” has been replaced by “reference period” throughout the text to ensure consistency in terminology. 2005 is the last year in the historical runs for the EURO-CORDEX CMIP5 ensemble, which is why we have taken the period 1976-2005 as a reference.*
- Line 111: The meteorological forcing is on a 8 km grid, but small catchments up to 64 km² (line 206) are included. Given that the effective resolution of the climate model is > 12 km and that of SAFRAN also > 8 km, which is more reflected by the underlying station distribution, I have the impression that too small catchments are included in the study. Can you comment on this?
- Table 2: Can you include the model resolution in Table 2? Right now, I need to look in the supplement. How did you bridge the gap from 8 km resolution after bias correction to the much finer hydrological model resolution?
 - ⇒ *Information regarding the size of the catchment is provided in line 207, “Given the 8-km spatial resolution of the gridded climate projections that were used as forcings for hydrological modelling, points with a drainage area of less than 64 km² have been excluded”. We have added the following sentence at the end of the section 4.2 entitled “Surface hydrographic network and actual simulated points”: “The minimum size of simulated catchments is 64 km², in accordance with the spatial resolution of climate projections” (§ 4.2 L238-239).*
- Fig 5: how do the results differ spatially?
 - ⇒ *Additional calculations will be presented and commented (L388-411).*
- 1 shows differences in bias correcting methods. This should be mentioned. As they are quite substantial, I would also like to see (despite condensing results) a graph showing not only a median change over all simulations (Fig 6), but also split into bias correction methods. Figure 5 would also be interesting to see and discussed for the bias corrections separately.
 - ⇒ *We have produced two graphs with the results of the two BCs (Figure 5). It should be noted that not all HMs used the climate projections obtained with CDF-t as inputs. The comparison will therefore not be complete (it will only be based on the GRSD, SMASH and MORDOR-SD models available at national level, see Table 2). It should be also noted that the uncertainty due to BCs in the Explore2 climate projections was found to be low, as the BCs applied are two similar approaches based on quantile mapping.*
- Table 3 (I refer to the second Table 3): Can you include formulas for the metrics (maybe in the assets or provide code)?
 - ⇒ *The formulas will be detailed in the supplementary materials (see Table S1).*
- Figure 3: some metrics are not mentioned in the text (aCDC, alphaQA). Consider to put them in value or delete the figure.
 - ⇒ *These maps have been commented (§ 5.2.1 L332-334).*
- Line 259 & 374: What do the authors mean with “historical runs”, I assume hydrological runs forced with climate model data from the period xxx to xxx?
 - ⇒ *To clarify the text, we will add the following sentence “Second, the entire modelling chain from GCMs to HMs was examined to assess how well the HMs forced by historical runs (i.e. bias-corrected projections for the reference period 1976-2005) can reproduce the various facets of the hydrological regime.” in § 5.1 L261-264.*

- Line 389: term “expected” is misleading as this metric can take on any value.
⇒ *We suggest this new sentence: “The criterion RRH should take a value between -1 and +1 if no major additional bias appears with the use of bias-corrected climate data” in § 5.2.3 L376-378.*
- Line 392: Which are the 17 runs? The owns with RCP8.5 only?
⇒ *A set of 17 pairs GCM/RCM were used to build the climate projections. Historical runs are common to the projections based on the same pair GCM/RCM: “A set 17 GCM/RCM pairs was finally retained to build the climate projections, with common historical runs (1976-2005) between projections based on the same GCM/RCM pair.” (§ 3 L158-161).*
- Line 438: title of section is misleading as also results mid of century are presented
⇒ *The title of § 6 is now: “Summary of the projected changes for mean river flows and groundwater recharge”.*
- Lines 439-440: Starting a results paragraph with saying that parts of the results are presented elsewhere reads strange.
⇒ *This sentence has be moved to the next paragraph.*
- Line 481: What do the authors mean with changes are limited when referring to Fig. 7? I can see substantial changes from ranging from minus to plus 50% covering most and parts of France, respectively?
⇒ *The changes in winter are limited compared to those projected for summer (see Table 6). Using the term “limited” is not relevant. We have removed this sentence, as it does not provide any additional information (details are provided later in the paragraph).*
- In whole section 6: I would recommend not to use the wording “significant change” as no statistical test was applied
⇒ *The reviewer is absolutely right. We have used “major changes” instead.*
- Line 504: The authors say that the most significant changes are in the summer. Looking a Fig. 7 RCP8.5 I do not see a large difference to fall.
⇒ *The decreases in summer are slightly greater than those projected in autumn (a difference of a few %). This small difference suggests that the changes observed in summer and autumn should be described in similar terms.*
- Line 514: North-south gradient. Please be a bit more specific to which season and which variable as I do not agree on the sentence “most maps show a north-south gradient”.
- Line 514: “This gradient is already present in the ensemble of climate projections”. Please add that this was not shown in the manuscript or put a figure in the Supplement.
⇒ *We have chosen the same colour palette for all maps, firstly adapted to display the changes predicted for the end of the century in the RCP8.5 scenario. This highlights this north-south gradient at the end of the century, but not necessarily for other scenarios and time slices. Spatial contrasts will be detailed using results obtained for two basins (one in the north and one in the south). Furthermore, in order to better understand the reasons for spatial contrasts, maps illustrating changes in precipitation will be added to the supplementary materials.*
- Line 546: Please mention that with the storylines only climate model uncertainty is presented (and maybe only a part), while the other sources are presented elsewhere.

⇒ *We have added this comment in § 3 L169-171 (“to illustrate uncertainties due to climate models for the end of the century, we developed a storyline approach introduced by Shepherd et al. (2018)”) and in the discussion.*

- 9: Do you want to discuss why the yellow story line (which should be a moderate one) is more extreme than the median changes presented in Figure 7? This is pretty obvious by design, but I recommend the authors to do not leave these kind of thoughts all readers.
⇒ *The comparison is not straightforward between Figure 7 and Figure 9, as the colour palette is not the same. WE have added a comment in the caption of Figure 9 (L 557).*
- Fig. 10 is not mentioned in the text.
- Fig. 10: I would suggest plotting also all individual projection.s with moving averages. The annual variability is not the point here to my opinion.
⇒ *Fig. 10 has been deleted in the main document and has been moved to the supplementary materials (Fig. S4) to support, in the text, the pattern of changes (north-south gradient).*
- Line 610 points to an outdated data set, there is a new doi provided
⇒ *The sentence has been modified: “The Explore2 dataset is associated with the following digital object identifier <https://doi.org/10.57745/YHMBHC>”*

RC2: 'Comment on egusphere-2025-1788', Anonymous Referee #2, 16 Oct 2025 reply

Sauquet et al. have submitted a manuscript detailing the development of a hydrological climate projection dataset for France, named Explore2. This dataset builds on a previous dataset, developed in the early 2000s. The work contains the collection of 36 EURO-CORDEX simulations, bias-correction and subsequent hydrological modelling.

When reading my review, it is relevant to know that I have a climate scientific background, rather than a hydrological background. As such, I have mostly focussed on the climate projections part, the logic of design decisions and clarity of text/concepts.

In general, I think the project and hence the manuscript are of high relevance. High quality climate and hydrological projections are of high value to any society that aims to adapt to future climate change. As such, I see potential for the submitted manuscript. However, in my opinion the writing of the manuscript has not been done very well at all. There is bad flow, important details are only mentioned as a by-the-way, and I kept having to go up and down to figure out how many and which simulations/datasets were being used at what point. In the discussion of the results I frequently was confused as to what the aim of the manuscript was, given many references to other papers showing the results.

I therefore recommend the editor to request major revisions before considering publishing in HESS.

Major comments

CLIMATE SERVICES

Explore2 is a classic 'climate service', I suggest the authors add some literature on good practices of climate services in their paper/introduction. Part of the discussion from line 416 onwards might suit in this part of the introduction. Also, please clarify the relationship between Explore2 hydrological projections and existing meteorological projections from MeteoFrance. Are the two related and how, if not, how are users meant to combine or choose between the two? Hydrology is a classic cross-border problem, with for example shared river basins and shared aquifers.

CLIMATE SCIENCE BASIS

Section 2 - This section is vital in your Explore2 project/data stream and hence in consequent use of the data. Without understanding of the source of climate data, how they have been bias-corrected, and how they sit in the range of uncertainty projections from CMIP/CORDEX/etc. sources, I would argue the resulting data lose a large part of their value for users. The current description of data, data treatment and uncertainties are incomplete and lacking clarity. In my opinion, this needs to be addressed before publication in an international scientific journal like HESS can be considered.

The Explore2 project is a large-scale project including the production of both climate and hydrological projections. The focus in this paper is on hydrology, which is why the climate section is only presented briefly. In light of the comments, a specific data paper was recently submitted (Marson et al., The Explore2-2022 climate projections dataset for impact studies over France, submitted). More details on the development of the climate dataset can already be found in section 2.3 of (Corre et al., <https://doi.org/10.1016/j.cliser.2025.100553>, 2025).

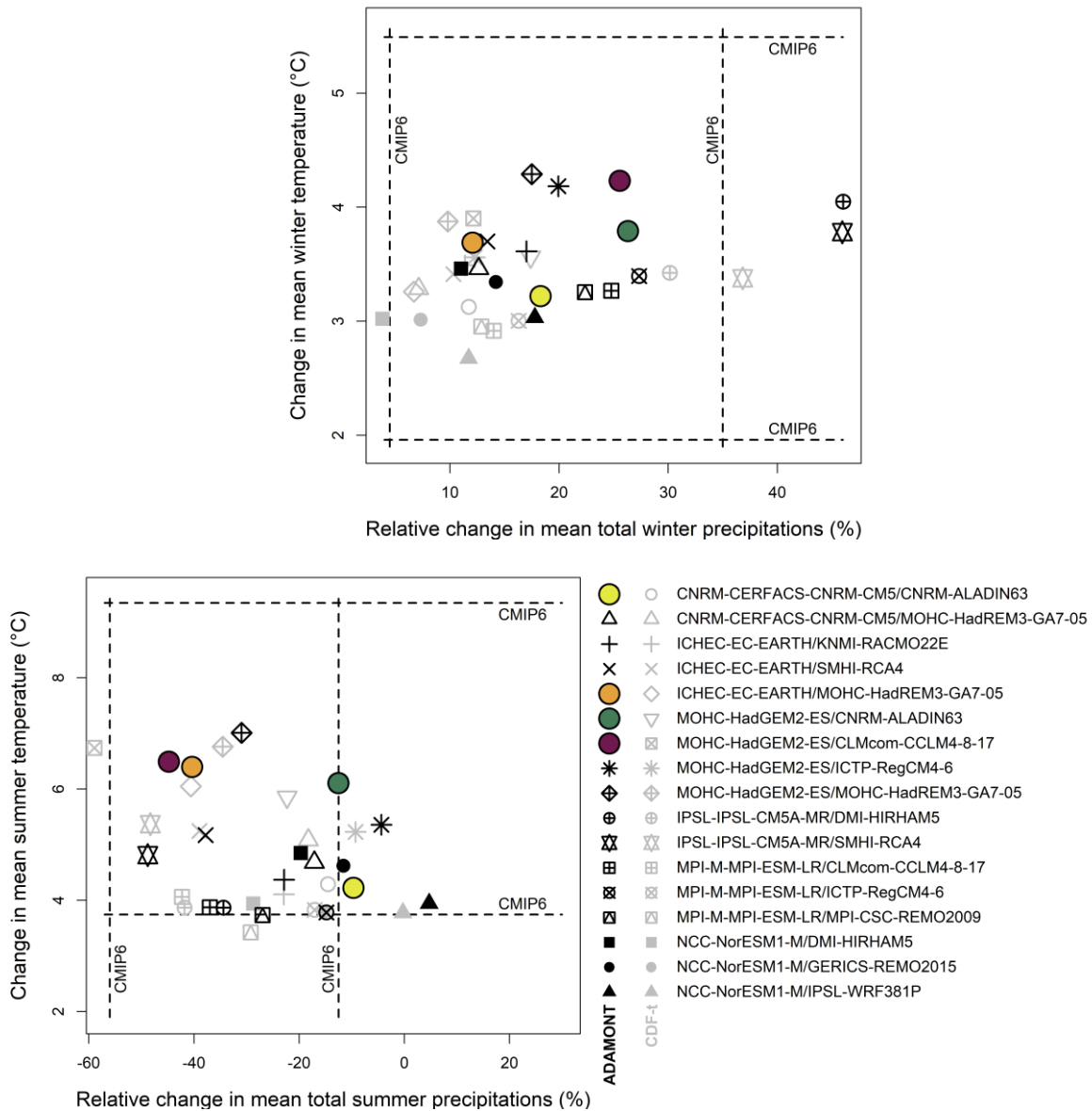
MANUSCRIPT

- Writing style is vague. Many sentences are unnecessarily long or round-about. Many sections are relevant but empty (multiple 2 sentence sections!) Please critically re-read and assess the different paragraphs and sentences to improve content and clarity.
⇒ *Multiple two sentence sub-sections has been deleted.*

Minor comments

The most important detail, for me, is which climate data underpin the Explore2 dataset and is not mentioned. The report that supposedly details it is written in French, Section 2.1 does not offer any information beyond a three-sentence standard introduction of EURO.CORDEX.

- Please include more details on your bias-correction methods. Why were these two methods chosen, are they uni- or multi-variate, how are they combined? The information about SAFRAN being the baseline reference should probably appear earlier, where you introduce the climate model data?
⇒ *Several bias-correction methods have been tested. Finally two were selected from the expertise (in particular, based on the realism of the rainfall fields). Both methods are univariate, we have added this information (“Both methods are univariate and applied independently to the outputs of each GCM/RCM pair” in § 3, L147-148). They are not combined, but applied independently to the outputs of the GCM/RCM pairs. A short description of SAFRAN has been given in the data section (§ 2, L112-118): “SAFRAN is a meteorological surface reanalysis available with a spatial resolution of 8 km and hourly temporal resolution, from 1 August 1958 to the present day (Vidal et al., 2010). This dataset was calculated using the Safran analysis system (Quintana-Seguí et al., 2008), which performs optimal interpolation between all surface observations available in the Météo-France database and the first estimate of the ERA-40 reanalysis (Uppala et al., 2005). A temporal interpolation step followed by a spatial interpolation step using 8 km resolution orography are applied in order to obtain daily data on a regular grid covering France, named grid SAFRAN hereafter. The meteorological reanalysis SAFRAN is used here to adjust the statistical biases of the climate simulations, and to calibrate and assess the performance of the HMs.”*
- Show me, rather than tell me, that you have assessed the compatibility with CMIP6. If you feel such figures are not the main message of this paper, you can add them to the supplement. For example, I know that CMIP6 models are much warmer than CMIP5 (higher climate sensitivity), how have you dealt with that?
⇒ *Consistency with CMIP6 ensemble: changes under RCP8.5 were compared to those projected by the CMIP6 ensemble under the SSP5-8.5 scenario for the end of the century and for the two contrasting seasons DJF and JJA. The objective of the comparison was to exclude projections out of the range of the CMIP6 ensemble for both temperature and precipitation changes in winter and summer. Please see Fig. 1 (modified version) and § 3, L149-156):*



Changes in seasonal precipitation and temperature (winter and summer) between the benchmark period 1976-2005 and the end of the century 2070-2099 under RCP8.5. Markers in grey are related to projections obtained with the CDF-t bias correction method while markers in black are related to those obtained with the other bias correction method ADAMONT. Dotted lines define the range of values consistent with projected changes in CMIP6 (calculations were made in 2022).

- I very much like the idea of offering a subset of ‘representative’ storylines for users. Your statements of ignorance and misuse are a bit harsh though, and without examples/references don’t sound very scientific. Furthermore, please add details on how the four storylines were chosen. In Table 1 if the median is shown in brackets, do I assume right that the other values are means (add to caption please), in Figure 1 what is the difference between grey and black markers?

⇒ *It is important to recall the concept of storyline introduced by (Shepherd et al., 10.1007/s10584-018-2317-9, 2018 (§ 3, L169-181)). We have developed the concept in few lines. A hierarchical clustering method based on mean changes of seasonal temperature and precipitation was used to classify the 17 projections obtained under the scenario*

RCP8.5 and on projection from each group was selected (expertise). The median and extreme values (minimum and maximum in brackets) for the Explore2 dataset for RCP8.5 are shown in the last column. Markers in grey are related to projections obtained with the CDF-t bias correction method while markers in black are related to those obtained with the other bias correction method ADAMONT.

- “The availability of an updated and state-of-the-art ensemble of climate projections since the publication of Explore2070 conclusions, and the increased needs of users underpinned the idea of a new project called Explore2.” → “A new project, Explore2, was initiated to take advantage of the availability of updated climate projections (CMIP6⁵) and to address the increased needs of users since Explore2070.”
- “This section describes the climate projection dataset used for deriving the Explore2 dataset through the use of hydrological models (HMs).” → “This section describes the climate projection dataset used for deriving the Explore2.”
- “Projections from the 5th Assessment Report instead from the 6th Assessment Report were selected for Explore2 as they were, at the moment they were retrieved, the only large set of high-resolution projections over France.” → very unclear sentence?
⇒ *These sentences have been corrected in the next version following your recommendations. Thank you.*
- Very very short sections: 2.1, 2.2, 3.2, 3.3
⇒ *These subsections have disappeared in this new version.*
- Line 60 - “large noise added by internal variability”: whilst I agree that internal variability adds noise when one is trying to quantify the forced climate response, for adaptation design I wouldn’t call it a source of noise. Instead, good adaptation practise includes preparing for the range of possible climate outcomes, which are determined by the combination of internal variability and forced climate change.
⇒ *I completely agree with the reviewer. The term ‘noise’ is not relevant. It has never been used in discussions with stakeholders. Thank you for pointing this out in the text.*
- Line 64 - can you include references (scientific or other, policy documents, etc.) that show how Explore 2070 has contributed to developing water management strategies?
⇒ *Unfortunately there is only few studies that used Explore2070, compared to the use of Explore2. Four years after the end of the project, Carroget et al. (<https://doi.org/10.14758/SET-REVUE.2017.22.02>, 2017) identified only few regional studies that used data from Explore2070. The project helped demonstrate the impact of climate change (illustrations from the Explore2070 can be found in the work of the committees of the Senate, e.g. <https://www.senat.fr/rap/r18-511/r18-511.html>) without necessarily leading to adaptation measures. Two publications can be found (in French). Lhuissier et al. (2016). Garonne 2050: a prospective study. <https://doi.org/10.1051/lhb/2016057>, and Goulard et al. (2019). The adaptation plan to climate change in the Adour-Garonne basin: planning to reduce territories’ vulnerability, <https://doi.org/10.4000/dynenviron.5320>. This was detailed in the end of § 1.1 (.*
- Line 68 - “The availability of an updated and state-of-the-art ensemble of climate projections” add a reference to which you mean? CMIP6, EURO-CORDEX, other?

- ⇒ *At the beginning of the project, the most comprehensive ensemble of regional climate projections was the EURO-CORDEX CMIP5 ensemble. First results from CMIP6 were available and were used to constrain the Explore2 ensemble of climate projections (see here above). See discussion in § 3 (L53-61).*
- Line 71 - what does 'two years of maturing' mean?
 - ⇒ *We needed two years to build the project including finding financial support, preparing the scientific program.*
 - Section 1.2 - I'm not sure it is relevant for the paper to understand your work package structure. Instead I suggest you add more information on the specific requests/boundary conditions that were set that Explore2 should adhere to. Interesting questions: how did you learn the user needs, was there any co-production process set up, did you evaluate HOW the Explore2070 data was used, and what users missed from there (apart from time horizons and spatial detail)?
 - ⇒ *Addressing the questions on link between scientists and end-users are highly important. As mentioned, we cannot develop this section: this will lead to a too long paper. A specific companion paper is in preparation. However, we will add few words on the use of Explore2070 (see above, § 3 L53-61).*
 - Line 159 - states that there are 12 models. Why are there only 9 models in Table 2? Or should I add the 4 models in Table 3? That makes 13 models? How have these models been selected?
 - ⇒ *The Explore2 project made use of 13 hydrological models (HMs) used in combination with the ensemble of climate projections described above. They include nine surface hydrological models, three groundwater hydrological models, and one empirical model used to calculate the potential groundwater recharge. These sentences were added at the beginning of § 4.1 (L191-197).*
 - Line 374 - these are important insights, maybe more suitable for the introduction or discussion than halfway your results. In the results I expect the proper/best way of assessing quality, not a methodological discussion.
 - ⇒ *The paper has been reorganised in that way. The evaluation strategy is the subject of a section dedicated (§ 5) apart from the section presenting the modelling chains.*
 - Line 439 - I'm confused, 3 other papers describe the results of the projections? I fully expected these to be the main result of this paper.
 - ⇒ *This article aims to present the Explore2 ensemble and the related modelling chains, how the HMs were evaluated, and the results concerning average river flow and groundwater recharge. This article was written to serve as the 'upstream' reference document. The three other articles detail the results on hydrological extremes (zero flow and floods, two aspects that require specific analyses and additional modelling, both papers are now available in HESS) and groundwater levels (still in discussion). It was not possible to present all the results in a single article.*
 - Line 546 - Another paper that supposedly describes what I expected in this paper?
 - ⇒ *This article addresses only the issue of uncertainties from the perspective of storylines. A more detailed analysis using an uncertainty partitioning method is presented in Evin et al. (<https://egusphere.copernicus.org/preprints/2025/egusphere-2025-2727/>), now*

accepted. This study assesses the uncertainties in this set using QUALYPSO, a method specifically designed to handle incomplete sets and to disentangle and quantify all sources of uncertainty, including that due to internal variability.

- Line 590 - More diversity is not always better, especially if one completely loses track of everything used (as in this manuscript). This sentence is a bit weird in that way, did the eFLaG and Australian projects have the same aims and scope? Where the assumptions/budgets/relevance similar? E.g. flood protection (studies) in the Netherlands gets far more attention than flood protection (studies) in, lets say, Austria.
 - ⇒ *More diversity is not always better, but controlled diversity is needed to assess uncertainty. The eFLaG project has an objective quite similar to that of Explore2 (water resources planning). Even if it was developed for industry purposes, “by providing a consistent dataset of future projections of these variables, eFLaG can potentially support a wide range of applications across other sectors. The predecessor, FFGWL, has been widely used within the water industry but has also found very wide application for diverse research purposes” (<https://essd.copernicus.org/articles/15/2391/2023>). For the Australian dataset, no specific application is mentioned in the paper by Zhang et al. (<https://doi.org/10.1016/j.jhydrol.2024.131286>, 2024). However, the research is part of a strategy to update knowledge and has received financial support from the Australian Government Murray–Darling Water and Environment Research Program (see acknowledgment). Indeed, assumptions, budgets and relevance for these initiatives (including Explore2) may not be similar. However, the Explore2 ensemble is the most developed one for assessing uncertainty and responding to local needs (future water availability in small catchment areas).*
- Line 597 - This paragraph I expected in the introduction, what is the policy reality where your dataset adds to. Now it is the closing statement, I’m left wondering how it relates and is useful for that reality.
 - ⇒ *In this last paragraph, we demonstrate the usefulness of the data produced by the Explore2 project, using a concrete example. The political will to adapt to global warming levels was announced after the launch of the Explore2 project. The climatic and hydrological contexts are based on the Explore2 data (see Corre et al., <https://doi.org/10.1016/j.cliser.2025.100553>, 2025 for the climatic context under different GWL).*