Reviewer #1

#RC1.1. This study explored the sources of uncertainty in different components of the model chain and investigated their contributions to two climate indicators and three hydrological indicators. The variation in performance was evaluated based on different regional characteristics. I consider the motivation of this paper very good, especially in the context of using ensembles for climate projection. The structure of the paper is well organized, and the presentation is good as well. However, I have a few concerns about the calculation methods that need to be resolved before the paper can be accepted.

We thank the reviewer for the overall positive comment. We understand these concerns and we will improve these aspects in the revised version. Most of these points have been addressed in previous articles, in particular the methodological paper Evin et al., 2019, https://doi.org/10.1175/JCLI-D-18-0606.1 and the application of the same methodology to a EUROCORDEX MME (Evin et al., 2021, https://doi.org/10.5194/esd-12-1543-2021). For this manuscript, the aim was 1/ to present the results of the uncertainty analysis for a very large ensemble of hydrological projections where uncertainty come from GCM, RCM, BAM and HM and 2/ to show how such a method allows to better understand where uncertainty models come from (from which model category first, but also from which individual models). We choose to explain the main assumptions of the method in Section 2 and provide the technical details in the Appendix. To answer the reviewer's concerns, we will provide a few additional details in Section 2 to clarify some technical points (e.g. estimation) and a new figure will be provided in the Appendix to illustrate the different steps of QUALYPSO.

Evin, Guillaume, Benoit Hingray, Juliette Blanchet, Nicolas Eckert, Samuel Morin, and Deborah Verfaillie. « Partitioning Uncertainty Components of an Incomplete Ensemble of Climate Projections Using Data Augmentation ». *Journal of Climate* 32, n° 8 (2019): 2423-40. https://doi.org/10.1175/JCLI-D-18-0606.1.

Evin, Guillaume, Samuel Somot, and Benoit Hingray. « Balanced Estimate and Uncertainty Assessment of European Climate Change Using the Large EURO-CORDEX Regional Climate Model Ensemble ». *Earth System Dynamics* 12, n° 4 (2021): 1543-69. https://doi.org/10.5194/esd-12-1543-2021.

#RC1.2. What is the purpose of applying cubic splines to the projection and what are the effects on the trend analysis (Line 583)? What is the meaning of the smooth trend denoted as CRi(t)? Please elaborate on the calculation method. Additionally, is the smoothing suitable for precipitation and hydrological indicators (especially max1D)?

The climate response of a simulation chain, denoted as CRi(t), corresponds to the long-term trend of the simulated projection (section 2.3.1). It is assumed to have a temporal variation that is inherently gradual and smooth. In this study, this long-term trend is estimated using a cubic spline model applied to the corresponding projection available for 1976-2099. As mentioned in the manuscript, other trend functions could be used to extract the climate response of each chain in QUALYPSO (linear trend, polynomial trend, etc.).

The calculation method is described in Section 2.3.2 and will be modified to clarify the reviewer's questions. Cubic smoothing splines are implemented by the function smooth.spline in R. For all indicators except temperature (e.g. seasonal precipitation, annual maxima of daily precipitation, and hydrological indicators), the inter-annual variability is relatively large compared to the long-term trend, so the smoothing parameter spar was set to 1.1 to reduce the model's flexibility. This prevents misattributing the low-frequency fluctuations caused by inter-annual variability to the climate response. For temperature, we apply a lower smoothing parameter value of 1 to provide more flexibility. The choice was defended in previous studies (Evin et al., 2021) and checked by visual inspection of the climate responses for this study. However, we agree that extracting the forced climate response can be difficult for some indicators, for example when they often reach a bound (e.g. zero for positive values) and/or when the interannual variability is large (as is the case for annual precipitation maxima). This point will be discussed in a new paragraph in the discussion.

#RC1.3. The authors may need to showcase some results from this step.

We thank the reviewer for this suggestion. An illustration of the climate responses obtained for one pixel and one catchment will be added to the manuscript in the Appendix.

#RC1.4. In the estimation of internal variability (Lines 593–600), why does the method first estimate Di(t) as the difference between the raw projection (Yi(t)) and CRi(t), rather than directly simulating variability from the raw projection over the target period? Does this step reduce or increase the internal variability? Based on the results, the internal variability is super large—could this be because the smoothing is not applicable?

As indicated in Section 2.3.1, The high- to mid-frequency fluctuations in the simulated projections result solely from interannual variability. Our approach assumes that it is reasonable to consider a trend model to estimate the climate response of a chain, and in turn fluctuations around, due to interannual variability (deviations from the climate response). Hingray et al. (2019) have shown that this assumption allows providing, for all uncertainty components, more precise estimates than estimates obtained with time-slice approaches. It does not reduce neither it increases the interannual variability but rather separate two components of the total uncertainty: variability of the climate responses and interannual variability. This point is also discussed in the introduction: "*Disentangling the climate response of a given chain from stochastic fluctuations caused by IV is key for a relevant uncertainty analysis. Estimating the climate response can be challenging, particularly for indicators such as precipitation, where IV is significant (Hingray et al., 2019). This difficulty arises because climate outputs blend the climate responses with chaotic fluctuations from IV, which propagate through all the subsequent models in the chain. If for a given GCM multiple members are available and used for subsequent simulations, the climate response of a modeling chain forced by this GCM can be estimated with the multi-member mean of the simulations, and IV can be estimated with the inter-member variability. However, many hydrological studies rely on single-member* *and time-slice GCM experiments. As a consequence, IV cannot be properly filtered out and, when they are not simply disregarded, stochastic fluctuations from IV are often attributed to GCM uncertainty (see, e.g., Bosshard et al., 2013; Vetter et al., 2017; Gangrade et al., 2020).*" It is true that interannual variability is often large in hydrological impact studies, because precipitation and hydrological indicators are highly variable from one year to the next.

Hingray, Benoit, Juliette Blanchet, Guillaume Evin, and Jean-Philippe Vidal. « Uncertainty Component Estimates in Transient Climate Projections ». *Climate Dynamics* 53, nº 5 (2019): 2501-16. https://doi.org/10.1007/s00382-019-04635-1.

#RC1.5. Please elaborate on the calculation of ESi(t), using one example (e.g., RCP, s). Why is a linear regression model applied, and how is it used (Line 608)?For Equation (A7), does this equation still work if incomplete or unbalanced ensembles are used? How are the effects of incomplete ensembles reflected in the results? Authors failed to explain this in detail since this is the second major question to be solved.

We thank the reviewer for this comment. The description of the estimation step will be extended and improved in the revised version. In short, the individual effects are estimated at once using the linear model A7 which describes a sum of additive terms (Samson et al., 2013). The estimation is implemented by the R function lm using least-squares (see l. 608 of the original manuscript) and standard recipes of numerical linear algebra (QR-decomposition).

Note that in a former application of QUALYPSO, the ANOVA was estimated with a Bayesian approach combined with a data augmentation technique (Evin et al. 2019). Estimates obtained with both approaches are almost identical. Both approaches provide unbiased estimates even when the ensemble is incomplete (see also section 8.1 in Evin et al., 2021). Compared to the regression approach, the Bayesian approach has the advantage of providing the uncertainty of these estimates. However, it is computationally demanding (roughly 100 times more than the regression approach). Estimates using the regression method are also more stable because they do not rely on the sampling of posterior distributions.

Evin, G.; Hingray, B.; Blanchet, J.; Eckert, N.; Morin, S.; Verfaillie, D. Partitioning Uncertainty Components of an Incomplete Ensemble of Climate Projections Using Data Augmentation. *J. Climate* **2019**, *32* (8), 2423–2440. https://doi.org/10.1175/JCLI-D-18-0606.1.

Evin, G.; Somot, S.; Hingray, B. Balanced Estimate and Uncertainty Assessment of European Climate Change Using the Large EURO-CORDEX Regional Climate Model Ensemble. *Earth System Dynamics* **2021**, *12* (4), 1543–1569. https://doi.org/10.5194/esd-12-1543-2021.

Sansom, Philip G., David B. Stephenson, Christopher A. T. Ferro, Giuseppe Zappa, et Len Shaffrey. Simple Uncertainty Frameworks for Selecting Weighting Schemes and Interpreting Multimodel Ensemble Climate Change Experiments. Journal of Climate. 15 juin 2013. https://doi.org/10.1175/JCLI-D-12-00462.1 .

#RC1.8. What is the difference between IV, RV, and FV? Should they use the same definition but with different superscripts/subscripts?

IV is the internal variability, i.e. the standard deviation of the fluctuations around the climate change responses (Eq. A8). RV is the variance of the residuals (l. 580) of the ANOVA model. It corresponds to the unexplained variance of the ANOVA model, i.e the variance of the climate changes responses that can not be explained by the sum of the main effects of the different models (GCM, RCM, HM) considered in the modelling chains. FVis the fraction of total uncertainty variance CCRV(t) resulting from each source of uncertainty (Eq. A10). For a given future time, one FV value is computed for each category of uncertainty source (i.e. for scenario uncertainty, GCM uncertainty, RCM uncertainty, HM uncertainty and RV) and those FV values sum to 1. They are different quantities, with different definitions. The illustration of QUALYPSO which will be added should clarify these differences.

#RC1.9. Does the selection of the time span length (i.e., 30 years in this study) affect the results, since a longer time span would likely lead to larger internal variability?

As indicated at l. 591-592, the climate change responses are taken as the absolute or relative differences of the climate responses for the center of the 30-year time period. As the climate response is estimated with a trend model (here a cubic spline), no time span is considered to estimate it. For the sake of simplicity, we refer to 30-year periods but the climate response of a given such period is the value of the trend model for the year at the center of this period. Consequently, the time span does not affect the climate change response. It neither affects the internal variability which is estimated from the annual deviations from the climate response (i.e. the long-term trend).