Community comment 1 Rasmus Benestadt

#CC1.1. I think this paper is very interesting and a welcome contribution. I also appreciate the opportunity to discuss some of the points made herein.

Thanks for this positive feedback as well as the open discussion on different points mentioned below. They have been taken into account in the revised version of the manuscript.

#CC1.2. One point raised is "Model uncertainty arises from model imperfections" which is important, but this paper neglects uncertainties connected with the downscaling approach because it fails to mention others then dynamical downscaling (aka regional climate models, abbreviated as 'RCMs'). There are also other ways of downscaling global climate models (GCMs) which are based on entirely different assumptions and come with different strengths and weaknesses. We expect them to produce similar results if they all are skillful, independently of each other. Hence, if dynamical and empirical-statistical downscaling give similar outlooks, then the results can be considered as being more robust. Therefore, I recommend that the paper includes some discussion on empirical-statistical downscaling in order to get a more complete picture on uncertainties associated with modelling.

We agree that it would be worth mentioning empirical-statistical downscaling approaches in the discussion, in particular in the subsection 5.4.1. which mention the limitations of the climate models. A paragraph will be added to discuss alternative downscaling approaches, including empirical-statistical methods and recent artificial intelligence approaches.

#CC1.3. The need of bias-adjustment also introduces uncertainties. It's in a fashion similar to 'sweeping the problem under the carpet', but also it assumes that the present biases are similar to those in a changed climate.

There is at least one example of downscaling precipitation statistics large multi-model CMIP ensembles that may be of relevance: https://doi.org/10.5194/hess-29-45-2025. However, this example focuses on downscaling daily precipitation statistics and may require an additional step using weather generators to produce time series needed as input for hydrological models. On the other hand, the downscaled precipitation statistics provides a rule-of-thumb estimate for number of days per year with heavy rainfall. The method described in this paper will provide a basis for studying the connection between climate change and water-born diarrhoea outbreak in the EU-SPRING project (https://www.springsproject.eu/).

One motivation for downscaling statistical properties (e.g. parameters of statistical distributions) is that statistical properties often are easier to predict/quantify than individual outcomes.

In some cases, climate internal variability (IV) actually provides some useful information about inter-annual variability and the range of plausible outcomes. For example, downscaled results of large ensembles provide a band of plausible temperatures in https://doi.org/10.1073/pnas.2503806122 that can be compared with historical temperatures, and such an evaluation reveals whether the downscaled results match the observed inter-annual variability. The mean of the model spread can be interpreted as the climate normal, whereas upper and lower limits represent hot and cold years. It is also interesting to note that the ensemble spread in some cases is close to being normally distributed.

We agree that statistical downscaling methods can be a valuable approach for producing statistical properties of meteorological variables under climate change. Applying weather generators to produce continuous time series based on these indices could be an option. Such approaches are actually under consideration in the context of the project Explore2 and are under development.

Large multi-model ensembles (with multiple runs) are obviously of interest to better estimate the expected mean and variability of present (and future) climates. Often, these large MMEs are not considered for hydrological impact studies because they are much too computational demanding. Note that as soon as transient climate projections are available, climate responses can be estimated from a single run, and the uncertainty assessment can be carried out even when multiple runs MMEs are not available, as is the case for most impact studies (Hingray et al. 2019). This is also illustrated in our work with QUALYPSO. Note that this "one member limitation issue", typically found in impact studies, is mentioned in the introduction.

#CC1.4. The statement "To our knowledge, the Explore2 MME is the largest ensemble of hydrological projections ever produced from regional climate experiments at the scale of a country" is probably true - see https://doi.org/10.5194/hess-29-45-2025 where MMEs were downscaled for SSP370, SSP126, SSP245, and SSP585, each with ~30 ensemble members (there are also unpublished results (work in progress) with downscaling total annual precipitation of 200-300 ensembles for each SSP).

When it comes to evaluation, it is not clear if the results are evaluated involving the complete chain of models. I.e. is the downscaling combined with hydrological modelling of GCM historical runs able to reproduce observed trends and inter-annual variability? Also, are the RCMs able to reproduce past variability and trends?

The evaluation issue of climate models is indeed a very critical issue. It is also a difficult one and requires it to be carefully addressed. This issue is out of scope of the manuscript and we could not give him the full place it should deserve. As mentioned in the manuscript, we refer to the studies led by the EUROCORDEX community for the evaluation of the climate simulations (e.g. Coppola et al., 2020). For the evaluation of the hydrological MME itself, we refer to https://doi.org/10.5194/egusphere-2025-1788.