Review of the research article titled "More intense heatwaves under drier conditions: a compound event analysis in the Adige River basin (Eastern Italian Alps)" and submitted to Hydrology and Earth System Science.

General Comments

This article presents a study on the impact of climate change on the intensity of compound drought-heatwave (CDHW) events affecting the Adige River Basin, a major basin in the Italian Alps. The authors use E-OBS and ERA5 data to characterize past meteorological conditions, plus two datasets from the Alpine Drought Observatory and the High-Resolution Pan-European Reanalysis to analyse the streamflow of the Adige. Then, they consider an ensemble of 25 EURO-CORDEX RCMs to assess the performance of climate models in reproducing historical trends in CDWH.

The CDWH are defined using daily maximum temperature (TX) and total precipitation (TP) transformed to obtain the SPI-6 index. After detecting all CDWH events since 1950 in ERA5, May 2022 is chosen as a case study, being the second most intense event after 2003, and the most intense in the last 15 years. To study trends in temperature and precipitation associated with this type of event, flow analogs based on 500 hPa geopotential height anomalies (Z500) are adopted. This allows to condition the analysis of the relevant meteorological parameters to the synoptic configuration that produced the specific event, and is a popular strategy to study (compound) meteorological and climatological extremes. For comparison, unconditional trends are also considered.

As it is often done in this type of study, the 1950-2022 reanalysis period was divided into a counterfactual (1951-1980) and a factual (1992-2021) periods, and changes in Z500, SPI-6, TX, TP are quantified between the two periods. Changes in quality of the analogs, annual analog frequency, seasonal distribution are also assessed, together with a simple evaluation of the influence from two large-scale natural variability modes. The analysis is then repeated using EURO-CORDEX simulations, extending the historical period to 2021 using RCP8.5 simulations.

The research question is interesting and coherent with current concerns about the impact of anthropogenic climate change on extreme events, and is well within the scope of HESS. The paper reads very well, all parts are well structured and clearly presented, and the title clearly states the scope of the study. Conclusions are clear and non trivial, and sustained by well described results, obtained with overall robust methodology. The results about the EURO-CORDEX performance (or lack thereof) are particularly interesting and useful, since great attention is being given to the future evolution of extreme events under global warming scenarios, and understanding how reliable models are for specific types of event is important. Also, figures are overall easy to read and well support the interpretation of results.

I found the paper to be overall almost ready for publication, however I would like to point to some minor issues that the authors should be able to address quite easily, detailed in the next section.

Very minor linguistic imprecisions are present here and there, I will only point out one or two that might affect text clarity, the rest should be addressed by the editors.

Specific Comments

Statistical testing

This comment is general and concerns the execution and presentation of statistical tests used to assess differences in the distribution of several variables.

Overall, I advise against reporting significance at different levels using starts. Although this is a common practice, classic statistical inference is structured around the idea of controlling the probability of type 1 error - the level of the test - and the latter should be specified before running the experiment. I suggest either fixing the level at alpha = 0.05 and reporting only significance at this level, or dropping the significance reporting and directly showing the p-value as in Figure 6. This particularly concerns Fig. 5 and 7a.

All differences throughout the paper are tested using a two-sample Cramer von Mises test. This test compares the integral difference between the empirical probability distribution functions of the two samples, and it is therefore adequate to assess overall changes in the distribution. However, most of these tests aim at testing a shift in the distribution. While in cases such as in Figure 5 the resulting significant differences are clearly due to a shift, in other occasions this is not the case. In particular, the analog quality in Fig. 6a results significantly different between the two periods, however this difference is very likely due to a change in variability, not a shift; the authors comment on this result stating that the circulation becomes more common, but I would say that this is not the case. To solve this, I suggest to use a different non-parametric test focused on the central tendency of the distribution, and not on the comparison of the entire distribution functions. A good alternative could be the Mann-Whitney U test.

Section 3.3.1

While there are several ways to use analogs to perform detection or attribution, the authors here follow the method proposed by Jezequel et al. (2018). In this case, analogs are used to perform a simple stochastic weather generation of events analogs to the May 2022 CDHW, which is not the only possible way to use analogs in general. The methodology is very briefly summarized at lines 186-190. I would suggest adding a few words to explain the technique, possibly using a bullet-point list of the steps, since this explanation could be a little hard to grasp for someone who is not used to the technique.

At the end of the section, the evaluation of natural variability is explained, and some limitations are stated. This method -comparing the distributions of the indices in correspondence of the analogs in the two periods - is currently being used not only in other studies, but also in rapid attribution projects (e.g. ClimaMeter). I would say that the main limitation is that, even if significant changes are (or not) observed, it is very difficult to make any statement about the causal relationship or even the correlation between this difference and those observed in the meteorological variables, especially for the AMO, that has a very long typical period. I suggest explicitly cautioning the reader about the fact that looking at teleconnection indices conditional to the circulation is not the same as looking at changes in circulation or impact variables caused by these teleconnections.

Section 4.2

Line 255, Figure 4: a clear trend in Z500 is found. However, as also stated later in the article, thermal expansion due to global warming directly causes an increase in geopotential heights. This does not equate observing a change in the circulation, since this is a global trend; Z500 should be detrended before performing this analysis, with previous studies suggesting to use a cubic rather than linear trend specification.

Minor comments

Line 20: the factual period is stated to be 1991-2020, while in the rest of the paper is 1992-2021.

Lines 112-113: "*...intense precipitations in autumn are mainly driven by cyclonic storms, and in spring/summer are due to snow melt processes, leading to a pluvial regime with two streamflow peaks…*" I might be missing something as I am not particularly expert in hydrology, but this sentence seems to state that snow melting is the cause of intense precipitation in the warm season. The authors probably mean that the two streamflow peaks are respectively due to strong cyclonic precipitation in autumn and snow melt processes in spring/summer, this sentence should be corrected.

Line 247: "*...we randomly reconstructed the atmospheric configurations […] based on the closest 20 analogs…*" I would rather say "we reconstruct the atmospheric configurations [...] using a stochastic simulation based on random sampling from the closest analogs". While there is a random component, most of the work is done by the analogs selection which is deterministic.