Comment on nhess-2021-121
Anonymous Referee #3

The manuscript evaluates the ability of 9 CORDEX RCMs to simulate meteorological (i.e., precipitation) and hydrological (i.e., streamflow) variables, as well as drought statistics, in the Cenajo basin (Southern Spain). The best RCMS are then used to generate future scenarios.

General comments

The manuscript is interesting and within the scope of the journal. From a methodological viewpoint, some relevant details need to be specified to understand the validity of the proposed approach, with special reference to drought analysis. Some sections should be re-organized. Also, the language must be improved in some parts.

Major comments

The title of the manuscript is wordy and redundant. Please rephrase.

LL 56-57: The authors state that “In literature few works analyze the reliability of RCMs considering meteorological droughts.” Please add references to previous studies on this topic and highlight the main differences with your study. In particular, the manuscript would benefit from a comparison with a recent study by Peres et al. (https://doi.org/10.5194/nhess-20-3057-2020), dealing with a statistical methodological framework to assess the skill of the EURO-CORDEX RCMs to simulate historic climate (temperature and precipitation) and drought characteristics (duration, accumulated deficit, intensity, and return period), at seasonal and annual timescales, in Southern Italy.

L 148: The authors have performed a lumped analysis in the Cenajo basin. To this end, they have to specify:

- If the reference grids of both the historical data and the CORDEX simulations are equivalent; If not, how do they pair the information from the two grids?
- how many grid cells fall within the Cenajo basin;
If the gridded historical and simulated precipitation data are spatially aggregated at the basin scale level and how.

The authors apply the SPI for meteorological drought analysis. However, it is not clear which time scale is used to aggregate monthly precipitation (1, 2, 3 months?) and which probability distribution is fitted to such data (gamma distribution?) for SPI computation. The authors should be aware that if they simply calculate the standard normal values corresponding to the differences of monthly precipitation data and the related monthly means, divided by the related monthly standard deviations, they do not obtain SPI, but another index known as the Standardized Rainfall Anomaly (Jones and Hulme, 1996), which is equal to the SPI only if aggregated precipitation data are normally distributed.

Moreover, once that the SPI series is computed by using different threshold values, drought characteristics such as frequency, length, magnitude, and intensity are determined. The authors do not clarify how these characteristics are computed. Nonetheless, I believe that, for instance, drought magnitude for drought events longer than one month has been computed as the sum of SPI values over the length. Is it correct? If so, the approach is misleading since a SPI value already quantifies the magnitude of a dry or a wet period occurs during the considered aggregation period.

The number of drought events identified for each considered threshold should be indicated in a table, together with the mean values of the corresponding characteristics. I am afraid that for the control scenario, very few droughts are identified for threshold values corresponding to severe and extremely dry conditions. Thus, I wonder how fair could be the comparison between observations and simulations? In addition, if the analysis is lumped (i.e., a single series for the whole basin is considered for each variable), it would be interesting to ascertain whether the drought statistics evaluated on RCM simulations correspond to the same drought events identified on the historical series.

Finally, bias correction through quantile mapping applied to SPI (if precipitation) or to SSI (if streamflow) series is a little confusing since these series are standard normal distributed by definition, therefore I do not expect big differences between the historical series and the control simulation series, unless due to sampling variability. Please clarify this point and explain the results illustrated in Figures 7 and 8.

Minor comments

L 10: Hydrological impacts of what? Maybe, change with “hydrological response“.

LL 10-12: “It assumes that ... when they provide better approximation to the historical basic and drought statistics.” This sentence is rather unclear and must be rephrased.

LL 18-20: In the last sentence there is no reference to the future scenarios of hydrological droughts.

LL 103 and 114: The term "goodness of fit" is usually applied to describe how well a statistical model (e.g., a probability distribution) fits a set of observations. I am not sure it is appropriate for RCM simulations.

L 162 and L 192: Sections 4.2 and 4.3 have the same title. Merge the two sections.

L 219: “the threshold of ”-“ 1.7 of SPI (considered to define extreme droughts ...)”.

Usually, -2 is used for extreme droughts.