July 16th, 2021

Dear Reviewer,

We are sending you the response to the referee comments (**nhess-2021-121-RC3**) of the manuscript -ref: **NHESS-2021-121-** and entitled "Do climate models that better approximate local meteorology improve the assessment of hydrological responses? Analyses of basic and drought statistics" by Antonio-Juan Collados-Lara, Juan-de-Dios Gómez-Gómez, David Pulido-Velazquez, and Eulogio Pardo-Igúzquiza.

We would like to express our sincere gratitude for your in-depth revision that will unquestionably help us to improve the manuscript. We have taken into account all the comments and we have provided response to them.

Thank you very much for your time and consideration.

Yours sincerely,

Antonio-Juan Collados-Lara, Juan-de-Dios Gómez-Gómez, David Pulido-Velazquez, and Eulogio Pardo-Igúzquiza.
Authors

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.

We thank the Reviewer for recognizing the interest and suitability of the paper for Natural Hazards and Earth System Science.

Major comments

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

We modified the title:

Do climate models that better approximate local meteorology improve the assessment of hydrological responses? Analyses of basic and drought statistics

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.

The main novelty is that in this work we also analyse the propagation of meteorological droughts to hydrological droughts. We have clarified it within the new version of the manuscript:

In literature few works analyse the reliability of RCMs considering meteorological droughts (Peres et al., 2020; Aryal and Zhu, 2021). In this work we also analyse the propagation of meteorological droughts to hydrological droughts. As far as we know, there are not studies that analyse if climate models that provide the best approximations of the local historical meteorology provide also better assessments of the hydrological impacts.

Added references:

Aryal, Y., Zhu, J.: Evaluating the performance of regional climate models to simulate the US drought and its connection with El Nino Southern Oscillation, Theor Appl Climatol, https://doi.org/10.1007/s00704-021-03704-y, 2021.

Peres, D. J., Senatore, A., Nanni, P., Cancelliere, A., Mendicino, G., and Bonaccorso, B.: Evaluation of EURO-CORDEX (Coordinated Regional Climate Downscaling Experiment for the Euro-Mediterranean area) historical simulations by high-quality observational datasets in southern Italy: Insights on drought assessment, Nat. Hazards Earth Syst. Sci., https://doi.org/10.5194/nhess-20-3057-2020, 2020.

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.

We clarified these points in the new version:

We used historical climatic data (precipitation and temperature) provided by Spain02 v2 dataset (Herrera et al., 2012) for the period 1972-2001. In this work we performed a lumped analysis in the Cenajo basin. The RCMs were retrieved from the CORDEX project (CORDEX PROJECT, 2013), with a spatial resolution of 0.11º (approximately 12.5 km). Note that Spain02 dataset uses the same reference grids than CORDEX project. The most pessimistic emission scenario (RCP8.5) for the future horizon 2071-2100 was selected for the future projections. For this scenario we analysed nine RCMs corresponding to four different General Circulation Models (GCMs) (see Table1). In our case study 33 cells of the grid mesh fall within the basin. The historical and simulated (from RCMs) precipitation and temperature were aggregated at basin scale considering a weighted average value according to the area of each grid mesh inside the basin. We also used official monthly natural streamflow data within the Cenajo basin for the historical period 1972-2001 (adopted as reference). These data were taken from the available information coming from the Spanish Ministry for the Ecological Transition and the Demographic Challenge.

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.

We calculated SPI using an aggregation time of 12 months and the calculation method is comprised of a transformation of one frequency distribution (gamma) to another standardized frequency distribution (normal). The same procedure was applied to streamflow to obtain the SSI. We have clarified it in the new version:

The meteorological and hydrological drought analysis was developed by applying the Standard Precipitation index (SPI) (Bonaccorso et al., 2003; Livada and Assimakopoulos, 2007) and Standard Streamflow index (SSI) (Salimi et al., 2021), respectively. They were estimated for periods of aggregation equal to 12 months. The calculation method requires the transformation of a gamma frequency distribution function to a normal standardized frequency distribution function. The statistics of the SPI/SSI series are obtained by applying the run theory (González and Valdés, 2006; Mishra et al., 2009) for different SPI/SSI thresholds from the lower SPI/SSI to 0.

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.

We calculated the magnitude for each drought event as the sum of the SPI values over the length of this event but the magnitude associated to a given threshold was calculated as the average of the magnitude for all the droughts events for this thresholds. We have clarified it and the calculation of the others droughts statistics in the new version:

The frequency is defined as the number of droughts events for each SPI threshold. For each drought event, we assess its duration as the number of months that the SPI is below a given threshold, its magnitude as the summation of the SPI values for each month of the event, and its intensity as the minimum SPI value. For each threshold we estimate the mean duration, magnitude, and intensity as the mean values of the cited variables for all the drought events.

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.

The number of drought (frequency) is also represented in the figures together with mean length, mean magnitude, and mean intensity. The number of droughts of the control scenario (Fig.8a) is different depending on the RCM, but, after the correction (some of them are similar to the historical), all simulations have similar values of number of droughts (Fig. 8b).

On the other hand, the droughts events identified for the historical series do not correspond to the same events in the corrected control series. Note that RCM simulations do not simulate specific days or months. They simulate climate, providing plausible meteorological series for a specific climate conditions. The bias correction approach is also designed to obtain the climate change signal (the statistics of the series) but not to reproduce the historical series from the control simulations. For example we represented in Figure RC3-1 the months in the historical period with SPI values below zero, for the historical and corrected control simulation series, and the obtained events are different.

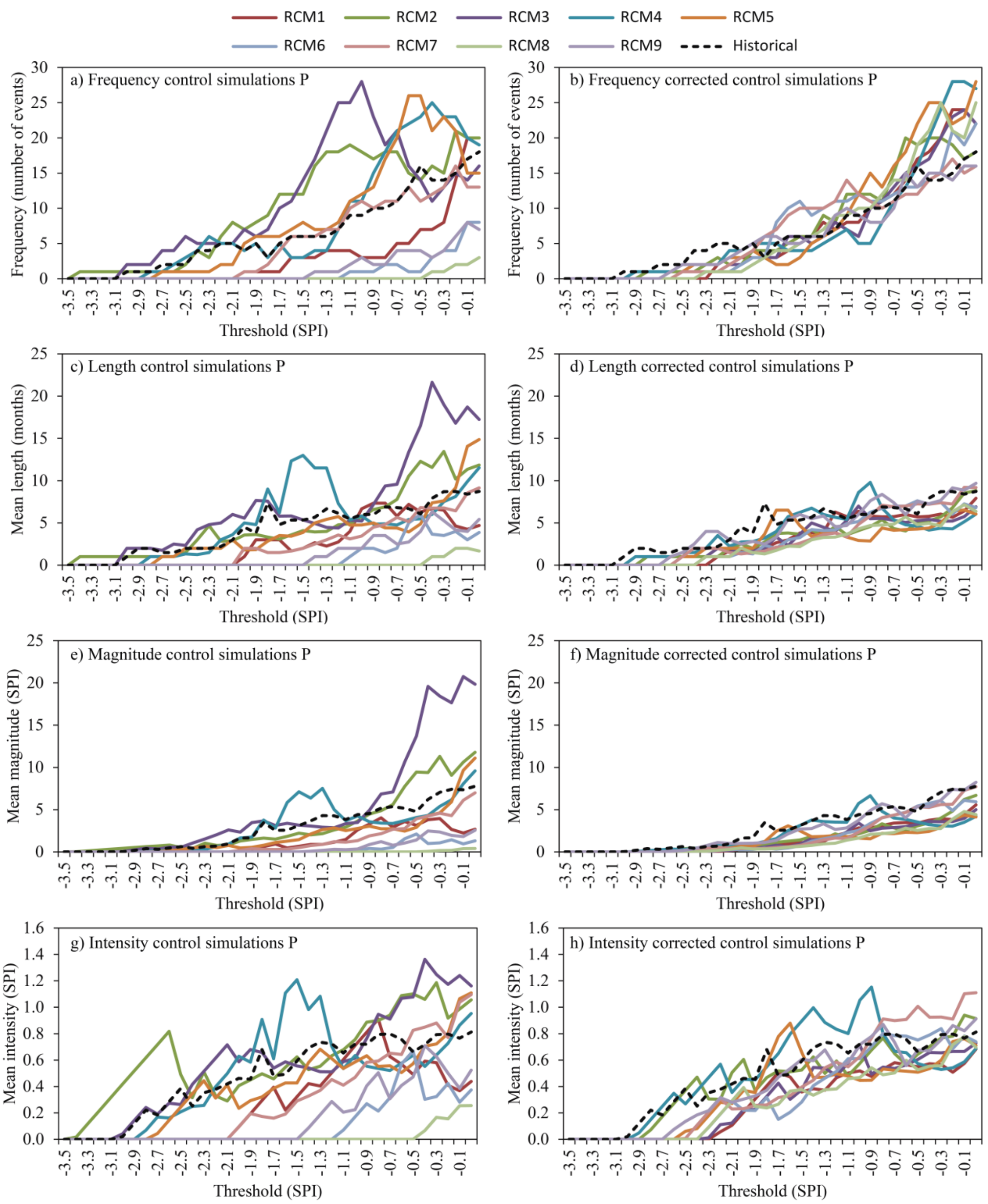

Figure 8: Drought statistics (frequency, length, magnitude and intensity) of the period (1972-2001) for the historical and control simulation series (left column) and historical and corrected control simulation series (right column) for precipitation (meteorological droughts).

Figure RC3-1. Months in which the SPI value is below zero for the historical and corrected control simulations series for the used RCMs.

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.

The corrected control series generated by applying a quantile mapping bias correction to the RCMS simulations show a very good fit with respect the historical series in terms of basic statistics. In terms of droughts (calculated from SPI), the bias correction approach clearly improves the fit of the RCM simulation series) to the historical series but the performance is lower than the one obtained for basic statistics. The left panel of the Figures represents the statistics before bias correction and the right panel after bias correction .The objective of this work is to classify RCMs according its capacity to reproduce historical basics and droughts statistic. Other bias correction technique should be explored to improve the performance for droughts statistics. We have introduced some changes in the new version of the manuscript in order to clarify it.

In the case of meteorological droughts (calculated from SPI) the bias correction approach clearly improves the fit of the RCM simulation series to the historical series for the four considered statistics (frequency, duration, magnitude and intensity). Note the differences between the left panel of Fig. 8 (control simulation and historical series) and right panel of Fig. 8 (corrected control simulation and historical series). For frequency the mean of SE for all the RCMs before the correction is 0.69 and after the correction is 0.23. For duration, magnitude and intensity these values are respectively 0.51 vs. 0.17, 0.88 vs. 0.30 and 0.38 vs. 0.13. In the same way, hydrological droughts were studied considering the SSI. Significant improvements are also observed for hydrological droughts (Fig. 9) after the bias correction procedure: frequency (mean SE of 0.63 vs. 0.34), duration (mean SE of 0.50 vs. 0.23), magnitude (mean SE of 0.83 vs. 0.51), and intensity (mean SE of 0.48 vs. 0.15). The left panel represents the droughts statistics of the historical and control series before applying the bias correction technique and the right one after a bias correction approach.

We also stated in the new section "5.1 Hypothesis assumed, limitations and future works" the necessity of exploring additional bias correction techniques to improve the performance for droughts statistics:

The corrected control simulation series obtained by using a quantile mapping bias correction presents a very good performance with respect the historical series in terms of basic statistics. In the case of droughts (calculated from SPI/SSI) the bias correction approach clearly improves the fit of the RCM simulation series to the historical series, but the performance is lower than for basic statistics. Other bias correction procedures should be explored to improve the performance for droughts statistics.

Minor comments

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

Done:

This work studies the benefit of using more reliable local climate scenarios to analyse hydrological responses.

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.

Done:

It assumes that Regional Climate Models (RCMs) simulations are more reliable when they provide better approximations to the historical basic and drought statistics after applying bias correction to them.

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

Done:

These two RCMs also predict higher changes in mean streamflow (-43.5 and -57.2 %) and hydrological droughts. The two RCMs also predict worrying changes in streamflow (-43.5 % and -57.2 %) and hydrological extreme droughts: frequency (from 3 to 11 and 8 events), length (8.3 to 15.4 and 29.6 months), magnitude (from 3.98 to 11.84 and 31.72 SSI), and intensity (0.63 to 0.90 and 1.52 SSI).

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.

We changed this term by performance along the manuscript.

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

It was an error. We changed the title of Section 4.3 to "4.3 Classification of RCMs"

L 219: "the threshold of "-" 1.7 of SPI (considered to define extreme droughts …)". Usually, -2 is used for extreme droughts.

We used the -1.7 SPI to identify extreme droughts because the Droughts Plan of the Segura River basin authority (where the Cenajo basin is located) proposes to use this value, but, in the figures we also show results in the different figures for all the thresholds.