July 16th, 2021

Dear Reviewer,

We are sending you the response to the referee comments (**nhess-2021-121-RC1**) 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

GENERAL COMMENTS

The authors present an assessment of the implications of bias correction methods for the assessment of the effect of climate change impacts on hydrological drought in a Mediterranean catchment. They applied a well-known bias correction method and chose to evaluate the performance of each RCM based not only on conventional statistics, but also on drought statistics. The authors discuss methodological issues related to the comparison of the RCM performance through the application of a rainfall-runoff model at the monthly time scale.

The topic is relevant for the audience of Natural Hazards and Earth System Science, the objectives are properly identified, the methodology for the analysis is adequate and the conclusions are relevant and correctly supported by the results and discussion. The overall organization of the manuscript is adequate, and it is clearly written. The analysis clearly shows the agreements and discrepancies between results obtained with different climatic forcings for the hydrologic model of choice. Therefore, I support publication of the work in Natural Hazards and Earth System Science.

We thank the Reviewer for recognizing the positive aspects of our manuscript, the relevance of the topic and its interest for Natural Hazards and Earth System Science readership.

SPECIFIC COMMENTS

I have several suggestions and comments, which I believe would improve the paper:

a) On section 4.2, the authors present their first assessment of discrepancy between historical observations and RCM control simulations. From Fig 4 and Fig 5, I gather that most models do a poor job at reproducing observed climate in the case study basin, particularly in seasonality of rainfall and temperature. I suggest adding a table with a comparison of mean annual values of precipitation and temperature to provide an objective comparison.

Following the reviewer suggestion we have added the suggested table to the manuscript and we have also made references to it within the text.

Table 2: Mean annual values of precipitation and temperature for the historical and the RCM simulations (and corrected RCM simulations) in the reference period (1972-2001).

	Mean annual precipitation (mm)	Mean annual corrected precipitation (mm)	Mean annual temperature (ºC)	Mean annual corrected temperature (ºC)
Historical	623.6	-	14.0	-
RCM1	700.5	623.5	10.4	14.0
RCM2	550.7	623.1	10.4	14.0
RCM3	503.6	623.3	13.2	14.0
RCM4	571.7	623.6	10.1	14.0
RCM5	588.7	623.3	8.5	14.0
RCM6	833.6	623.7	9.9	14.0

RCM7	683.0	623.1	9.6	14.0
RCM8	952.9	623.3	10.9	14.0
RCM9	826.1	623.5	9.5	14.0

b) The application of the quantile mapping technique is a critical step in the analysis. However, the authors do not provide much information on the procedure or the results while applied to the case study. There is a very brief introduction in the methodology section, with no details on how the original series are transformed. Regarding results, we can only see that bias for the three basic statistics has been eliminated. I think the authors should provide more information on the application of the technique to the case study and illustrate it with at least a figure showing the quantiles.

We have included more information and references about quantile mapping technique within the methodology section:

The statistical transformation was defined by a quantile mapping technique based on empirical quantiles. We used the open-source R package qmap (Gudmundsson et al., 2012). Quantile mapping with empirical quantiles uses a non-parametric transformation function. In this approach the empirical cumulative distribution functions (CDFs) are approximated using tables of empirical quantiles. It estimates values of the empirical CDFs of observed and simulated time series for regularly spaced quantiles to create the table that relates observed and simulated time series (Enayati et al., 2021). The values between them are approximated by using linear interpolation. Accordingly, it uses interpolations to adjust a datum with unavailable quantile values. For each month of the year we used its table of empirical quantiles. These tables, which are obtained by using the CDF of the observed and simulated values (from RCMs), are also used to correct the future simulation (from RCMs). If the RCM values are larger than the historical ones used to estimate the empirical CDF, the correction found for the highest quantile of the historical period is used (Gudmundsson et al., 2012).

Following the reviewer suggestion we have also included a figure showing the precipitation and temperature quantiles for the observed and control simulation series obtained with the RCM1 for each month of the year in the reference period (1972-2001):

These differences force us to apply the correction approach defined in section 2.1 for all the RCMs considered. It uses the CDF (quantiles) of the historical series and control series obtained from the RCMs simulations to perform the correction. The precipitation and temperature quantiles of the observed and control simulation series of RCM1 in the reference period are showed in Fig. 6. The same information was generated for all the RCMs simulations and used to correct the RCMs outputs.

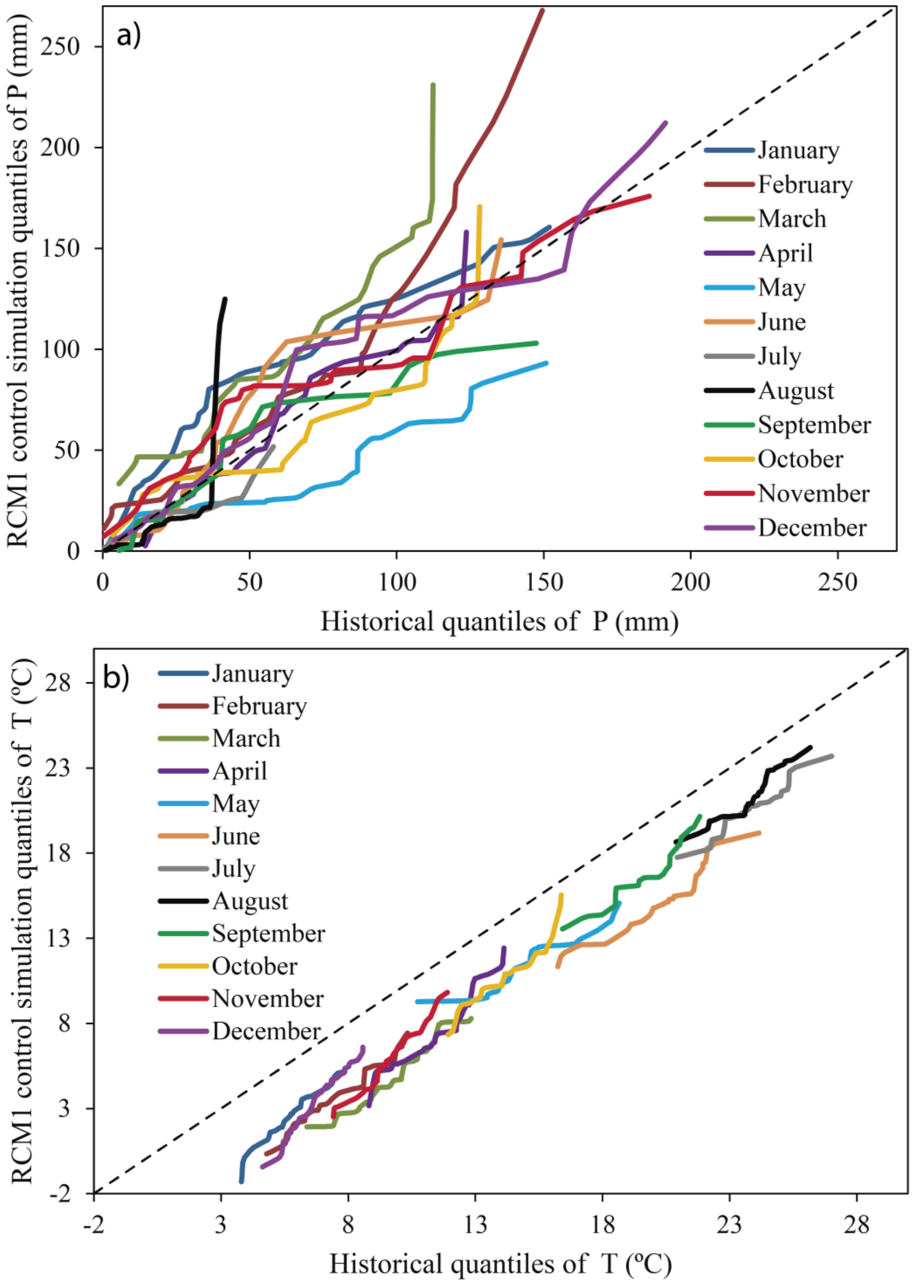

Figure 6: Precipitation and temperature quantiles of the observed and control series of the RCM1 simulations for each month of the year in the reference period (1972-2001).

c) The authors chose to use SPI as drought index to characterize precipitation, but they should state the aggregation time step chosen in the analysis. The descriptive statistics used later in the paper (frequency, duration, magnitude, and intensity) should be formally introduced.

We have included information about SPI aggregation time step and the used statistics of droughts:

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. 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 probability of occurrence of precipitation or streamflow for the SPI/SSI calculation, in the corrected control and future simulations, was obtained using the parameters calibrated from the observed series, in order to perform an appropriate comparison (Marcos-Garcia et al., 2017).

d) I was a bit confused by the classification procedure. If I understood correctly, the RCM are assigned penalty values from 1 to 10 according to their ranking in each of 7 statistics. The final classification is obtained by averaging of the penalization for all statistics. However, the index chosen is divided by a normalizing value to allow comparisons across statistics. Why not directly use the index values instead of the penalties based on the ranking, to account for the relative deviations shown by each model?

We propose these normalized values in order to give similar weight to all the statistics in the final classification. Note that the skew coefficient and droughts statistics have higher SE values. If we sum the SE values for all the statistics and we classify RCMs in accordance with it, the mean or standard deviation statistics will not influence in the final classification. It also allows us to define an index (SE) threshold below which the RCMs are not penalized. We have included it in the new version of the manuscript:

The penalization approach allows us to define an index (SE) threshold below which the RCMs are not penalized. It also allows us to give similar weight to all the statistics in the final classification. Note that the skew coefficient and droughts statistics have higher SE values. If we sum the SE values for all the statistics and we classify RCMs in accordance with it, the mean or standard deviation statistics will not influence in the final classification.

e) On section 4.3, line 195, the authors state that there is a "correlation" between the order classification of corrected RCMs for meteorology and hydrology. By looking at Figure 9, I am not sure of this and I am afraid I must disagree. Figure 9 shows a scatter plot of nine values. The fitted regression line for the nine points has an R2 of 0.34, which is very low to conclude that there is a correlation (what is the significance level?). Even the blue line, which corresponds to only to 4 points, has a very low R2, of only 0.46. Finally, the authors should refrain from plotting the regression line for the two points corresponding to classification order <2, which obviously renders a perfect fit because there are only two points. By looking at the figure, I can also see an opposite "correlation" for the 5 points corresponding to classification order >4. The fitted regression line would have a negative slope, contradicting the initial statement. I think *this discussion should be reformulated. We all agree that good bias correction would improve the agreement between climate models and observations, but the authors need to provide objective results to draw this conclusion, which, by the way, is a central part of their contribution. I suggest separating the analysis of conventional statistics and drought statistics, since the bias correction procedure is specifically focused on fitting the results of climate models to observations and therefore one can expect (as shown in Figure 4 and Figure 5), that the index values are very low. This does not necessarily have to be the case for drought statistics, which are linked to the tail of the distribution. Perhaps showing the scatter plots of the actual index values obtained with all models would illustrate better the comparison of performance for meteorological and hydrological drought.*

The reviewer is right, the figure is confusing. When the Classification order <2, it obviously renders a perfect fit because there are only two points. For these two points we wanted to highlight that the first and second best models for both analyses (meteorological and hydrological analyses) are the same RCMs (RCM9 and RCM2). The third "best" model for meteorology is the fifth in the hydrology assessment, and the forth in meteorology the third in hydrology assessment. The results show that the best models for meteorology provides also the best results for hydrology, but as the Reviewer pointed we can see an opposite "correlation", if the analysis is extended to other models that are not the best ones. This discussion can be supported by Table 3, and for this reason, in accordance with the reviewer comment, we propose to eliminate Figure 9 in the new version of the manuscript. We have also modified section 4.3 to clarify it:

The classification of RCMs (after the bias correction of the simulations) based on the approximation of the meteorological and hydrological statistics (basic and drought statistics) by applying the procedure described in section 2.3 is included in Table 3. The two best corrected RCMs for meteorology (RCM2 and RCM9) are also the best models for hydrological assessment (maintain the first and second position in both cases). Nevertheless, the third "best" model for meteorology is the fifth in hydrological assessment, and the forth in meteorology the third in the hydrological assessment. Although they are still in the group of the best approaches, it demonstrates that there is not a cause-effect relationship; a better meteorological approximation not always means a better hydrological assessments. We only demonstrated that, in our case study, the RCMs that provide the best approximations of the meteorology provide the best assessments of hydrological impacts.

TECHNICAL CORRECTIONS

From the formal standpoint, the paper is well written, correctly organized and adequately illustrated with tables and figures. I think the authors should rethink Figure 9 entirely.

We thank the Reviewer for recognizing the positive aspects of the paper. Figure 9 was deleted because it was confusing (as the Reviewer pointed) and our findings are properly supported by Table 2.

The authors should consider changing the term "asymmetry" coefficient for "skew" coefficient.

We changed it along the text and figures.

Page 4, line 119. I believe the normalizing value used in the denominator of equation 1 is useful for comparisons across statistics, not across RCMs, because the normalizing value (historical observations) is the same for all RCMs.

The Reviewer is right. We modified the sentence:

Note that this index is a mean squared error of the corrected control with respects the historical values. It is divided by the square of the mean historical value in order to make the results comparable for different statistics.

Figures 8 and 13. Please change SPI into SSI, since the plots refer to streamflow droughts.

Done.

Although I am not a native English speaker, I believe the following expression should be corrected:

On page 4, line 114, … applying the "following" error index.

Done:

We assessed the performance for each RCM in the reference period by applying the following error index (SE):

On page 4, line 119, … in order to make it comparable .

Done:

Note that this index is a mean squared error of the corrected control with respects the historical values. It is divided by the square of the mean historical value in order to make the results comparable for different statistics.

On page 6, line 174, … when "they" compared different statistical techniques.

Done:

It confirms the results obtained by Collados-Lara et al. (2018) when they compared different statistical correction techniques.