Dear Reviewer,

We are sending you the response to the referee comments (**nhess-2021-121-RC2**) 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

This study aims to provide further insights on the selection of Global Climate Models (GCMs) -Regional Climate Models (RCMs) combinations, according not only to their skills to reproduce the local climate during the selected historical period but also the local hydrology. Concretely, the authors calculate an error index for basic and drought statistics and use it to classify the GCMs-RCMs combinations according to their reliability for the assessment of meteorological and hydrological impacts. The selected methodology involves the bias correction of climate models' outputs through a quantile mapping (QM) approach based on empirical quantiles, the use of a lumped rainfall-runoff model to simulate monthly inflows from climate data and the use of standardized indices for drought characterization (namely the Standardized Precipitation Index (SPI) and the Standardized Streamflow Index (SSI)).

In my opinion, the paper addresses an important issue on the use of climate models' outputs for the assessment of climate change impacts at the basin scale. Besides, it is properly written and well presented.

We thank the Reviewer for recognizing the importance of the issue and positive aspects of the manuscript.

However, I miss a more critical approach to the potential shortcomings of the selected methodology, such as the underlying assumption of stationary bias, the impact of bias correction on the tails of the distribution (e.g. induced changes on the original climate change signal of the climate models), the pros and cons of pre-processing and postprocessing the variables derived from climatic ones with regard to bias correction (e.g. performance of the hydrological model simulations over the validation period) or the potential effects of neglecting the inter-variable dependence of climate variables (e.g use of univariate bias correction methods against multivariable ones or ignoring the role of temperature in drought onset) on the assessment of climate change impacts on water resources.

The reviewer is right; the shortcomings of the selected methodology should be stated. We included a new section in Discussion to point the limitations and future works related to the proposed approach:

5.1 Hypothesis assumed, limitations and future works

Although we have demonstrated the utility of the proposed approach to assess future impacts on meteorological and hydrological droughts, we want to highlight some hypothesis and limitations assumed and to identify potential future research aligned with this study:

- We have used a bias correction method based on the assumption of bias stationarity of climate model outputs. However, this assumption may not be valid to study some problems due to the significance of the influence of climate variability on them. Other approaches should be explored to take into account non-stationarity bias of RCMs simulations (e.g. Hui et al., 2020).
- We have applied the same bias correction procedure for all the range values in accordance with the climatic variable distribution function. We did not consider

the impact of bias correction techniques on the tails of the distribution, which could be important to analyse extremes (Volosciuk et al., 2017).

- In this work a univariate bias correction method is used. It does not consider the dependence between precipitation and temperature which could be explored in future assessments. Meyer et al. (2019) found that incorporating or ignoring inter-variable relationships between temperature and precipitation could impact the conclusions drawn in hydrological climate change impact studies in alpine catchments.
- The streamflow information available for this case study cannot be divided into two long-enough (e.g. 30 years) series representative of the climate/hydrology to perform explicitly a validation of the bias correction models (Chen et al., 2021). We assumed that the statistics of any long-enough periods remain invariant. In this case the calibration implicitly could be considered validated, due to the fact that the same results would be obtained under this hypothesis for any other period representatives of the climate/hydrology conditions.
- In our case study the influence of temperature was considered only in the hydrological assessment by using rainfall-runoff models. However other meteorological droughts indices that consider temperature could be included in the analysis [e.g. the Standardised Precipitation-Evapotranspiration Index (SPEI) (García-Valdecasas Ojeda et al., 2021)].
- 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.
- The proposed method has not been tested in other typologies of basin, as for example in Alpine basins where snowmelt component may have a significant influence on the results.

SPECIFIC COMMENTS

Lines 27-29: "For instance, we have Palmer Drought Severity Index (...)". I will also mention the Standardized Precipitation Evapotranspiration Index (SPEI, Vicente-Serrano et al., 2010).

Vicente-Serrano S.M., Santiago Beguería, Juan I. López-Moreno, (2010) A Multi-scalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index - SPEI. Journal of Climate 23: 1696-1718.

Done:

For instance, we have Palmer Drought Severity Index (PDSI) (Palmer, 1965), Crop Moisture Index (CMI) (Palmer, 1968), Standardized Precipitation Index (SPI) (McKee et al., 1993), Soil Moisture Drought Index (SMDI) (Hollinger et al., 1993), Vegetation Condition Index (VCI) (Liu and Kogan, 1996), Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010). Lines 91-92: "This is the reason that justifies the selection of quantile mapping (using empirical quantiles) for this study". Have the effects of inter-variable dependence been considered before selecting an univariate bias correction method? In the case of alpine catchments, Meyer et al. (2019) found that incorporating or ignoring inter-variable relationships between air temperature and precipitation data could impact the conclusions drawn in hydrological climate change impact studies.

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., Cannon, A. J. (2019). Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments. Hydrology and Earth System Sciences, 23, 3, 1339-1354, https://hess.copernicus.org/articles/23/1339/2019/

We agree with the reviewer. We did not consider dependence between precipitation and temperature, which may produce significant impacts in some basins, as alpine catchments. Our case study is not located in an alpine catchment, but we have also added it as a potential limitation of the methodology:

In this work a univariate bias correction method is used. It does not consider the dependence between precipitation and temperature which could be explored in future assessments. Meyer et al. (2019) found that incorporating or ignoring inter-variable relationships between temperature and precipitation could impact the conclusions drawn in hydrological climate change impact studies in alpine catchments.

Lines 105-109: "The meteorological drought analysis was developed by applying the Standard Precipitation index (SPI)". What about the role of temperature? As multiple authors have already pointed out, SPEI usually shows more severe increases in future drought events than those from SPI (e.g. García-Valdecasas Ojeda et al., 2021) and therefore I recommend to include it in the analysis. Which aggregation periods, statistical distributions and thresholds are considered for both the SPI and the SSI?

García-Valdecasas Ojeda, M., Gámiz-Fortis, S.R., Romero-Jiménez, E., Rosa-Cánovas, J.J., Yeste, P., Castro-Díez, Y., Esteban-Parra, M.J. (2021). Projected changes in the Iberian Peninsula drought characteristics, Science of The Total Environment, Volume 757, 143702, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2020.143702.

In our case study the temperature was considered to propagate climate impacts on hydrology by using the rainfall-runoff model. We also included the SPEI as a possible index to be studied in future works:

In our case study the influence of temperature was considered only in the hydrological assessment by using rainfall-runoff models. However other meteorological droughts indices that consider temperature could be included in the analysis [e.g. the Standardised Precipitation-Evapotranspiration Index (SPEI) (García-Valdecasas Ojeda et al., 2021)].

We have also specified within the new version of the manuscript the aggregation periods, statistical distributions and thresholds considered for SPI and SSI:

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).

Lines 136-146: In my opinion, the climate and hydrological regime of the Cenajo basin should be properly characterized in the Case study section.

Done:

The Cenajo basin has a Mediterranean climate. In the period 1972-2001, the mean annual precipitation was 623.6 mm and the mean temperature 14.0 °C. In the same period the mean annual streamflow was 443.6 Mm³. This is a critical area where climate change will exacerbate these problems by reducing the availability of resources and increasing irrigation requirements. It will also cause an increase in the magnitude and frequency of extreme events, such as droughts.

Line 149: "CORDEX project (2013)". Reference?

Added:

The RCMs were retrieved from the CORDEX project (CORDEX PROJECT, 2013), with a spatial resolution of 0.11° (approximately 12.5 km).

Lines 151-152: "We also used official monthly natural streamflow data within the Cenajo basin for the historical period 1972 -2001 (adopted as reference)". This reference period is not consistent with the calibration period of the rainfall-runoff model (October 1971 to September 2007, line 157).

Yes, the periods are different. We used the available historical information to calibrate the rainfall-runoff model and selected a 30-year period, which is usually used to perform the climatic change analysis. We have clarified it within the manuscript:

The rainfall-runoff model for the Cenajo basin was calibrated and validated using the available monthly climatic data (precipitation, temperature, and potential evapotranspiration) and streamflow data for the period October 1971 to September 2007. We split the period with available data in two to perform a calibration (from October 1971 to September 1989) and validation (October 1989 to September 2007) of the model. The performance of the model was assessed by using the Nash-Sutcliffe efficiency (NSE) coefficient, the correlation coefficient (R2), and the root mean squared error (RMSE). These statistics and the historical and simulated streamflow series are

showed in Fig. 3a. For the entire period (October 1971 to September 2007) the performance is also good (NSE = 0.94) and it is higher (NSE = 0.96) if we focus on the monthly mean within the mean year for the entire period (Fig. 3b). The model was used to propagate the impacts of climatic variables on streamflow between 1972 and 2001, a 30 year horizon, which is a period of time usually used in climate change analysis.

Line 153: "(...) Spanish Ministry of Agriculture, food and environment". The competences of this former ministry have been assumed by the current Ministry for the Ecological Transition and the Demographic Challenge.

Thank you. We updated it:

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.

Lines 156-161: What is the validation period? Goodness of fit for the validation period?

Thank you to the reviewer we realized that we did not explain properly this paragraph. We only presented results for the entire historical period. In the new version we included information about calibration and validation periods and its performance statistics. We also modified Figure 3:

The rainfall-runoff model for the Cenajo basin was calibrated and validated using the available monthly climatic data (precipitation, temperature, and potential evapotranspiration) and streamflow data for the period October 1971 to September 2007. We split the period with available data in two to perform a calibration (from October 1971 to September 1989) and validation (October 1989 to September 2007) of the model. The performance of the model was assessed by using the Nash-Sutcliffe efficiency (NSE) coefficient, the correlation coefficient (R2), and the root mean squared error (RMSE). These statistics and the historical and simulated streamflow series are showed in Fig. 3a. For the entire period (October 1971 to September 2007) the performance is also good (NSE = 0.94) and it is higher (NSE = 0.96) if we focus on the monthly mean within the mean year for the entire period (Fig. 3b). The model was used to propagate the impacts of climatic variables on streamflow between 1972 and 2001, a 30 year horizon, which is a period of time usually used in climate change analysis.



Figure 3: Historical and simulated monthly streamflow series in the Cenajo basin for the calibration period (October 1971 to September 1989) and validation period (October 1989 to September 2007) (a) and mean monthly values within the mean year of the entire period (October 1971 to September 2007) (b).



The reason could be that we are neglecting the inter-variable dependence of climate variables not considering the dependence between precipitation and temperature when the bias correction is applied. Therefore, some differences might appear in the streamflow that depend on the combined interaction of both variables. We have included it within the new version of the manuscript:

The fit of the corrected control simulation series of streamflow to the historical series is not as good as for precipitation and temperature, but a remarkable improvement is observed. The reason could be that we are neglecting the inter-variable dependence of climate variables not considering the dependence between precipitation and temperature when the bias correction is applied. Therefore, some differences might appear in the streamflow that depend on the combined interaction of both variables.

We have also commented it in the new subsection 5.1 Hypothesis assumed, limitations and future works (line xx-yy of the new version of the manuscript):

In this work a univariate bias correction method is used. It does not consider the dependence between precipitation and temperature which could be explored in future assessments. Meyer et al. (2019) found that incorporating or ignoring inter-variable relationships between temperature and precipitation could impact the conclusions drawn in hydrological climate change impact studies in alpine catchments.

What about the performance over the validation period? (e.g. see Chen et al., 2021).

Chen, J., Arsenault, R., Brissette, F. P., & Zhang, S. (2021). Climate change impact studies: Should we bias correct climate model outputs or post-process impact model outputs? Water Resources Research, 57, e2020WR028638. https://doi.org/10.1029/2020WR028638

The available streamflow information cannot be divided into two long-enough (e.g. 30 years) series representative of the climate/hydrology whose statistics are nearly invariant, we cannot perform, explicitly, a validation of the correction model. We assumed that the statistics of any long-enough periods remain invariant. In this case the calibration implicitly could be considered validated, due to the fact that the same results would be obtained under this hypothesis for any other period representatives of the climate/hydrology conditions. We stated it as a hypothesis and limitation assumed in our approach:

The streamflow information available for this case study cannot be divided into two long-enough (e.g. 30 years) series representative of the climate/hydrology to perform explicitly a validation of the bias correction models (Chen et al., 2021). We assumed that the statistics of any long-enough periods remain invariant. In this case the calibration implicitly could be considered validated, due to the fact that the same results would be obtained under this hypothesis for any other period representatives of the climate/hydrology conditions.

Line 189: "Note that in this case we refer to the Standard Streamflow Index (SSI)". This index should be properly defined previously (in the Methodology section), along with an appropriate reference.

Done:

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.

Lines 198-199: "Therefore, we demonstrated that RCMs that allow better approximations of the meteorology provide better assessments of hydrological impacts". Although it seems quite straightforward (as rainfall-runoff models require climatic variables as inputs), I think that this statement should be carefully discussed before generalizing it: would it hold true if basins with very different hydrological regimes were considered? (e.g. important groundwater or snowmelt components?).

The Reviewer is right, we have modified this paragraph:

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. 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.

We also pointed in the section "5.1 Hypothesis assumed, limitations and future works" new version the interest of consider basins with different hydrological regimes to test the proposed method:

The proposed method has not been tested in other typologies of basin, as for example in Alpine basins where snowmelt component may have a significant influence on the results.

Lines 208-209: "Both RCMs predicts a decrease of the variability in precipitation and an increase of the variability of temperature in the future". This is an interesting result, as precipitation variability is generally expected to increase in a climate change context (e.g. Pendergrass et al., 2017). Concretely, for the Mediterranean regions, Polade et al. (2017) concluded that a decrease in the frequency of daily precipitation events, combined with an increase in the amount of precipitation delivered in relatively rare heavy events, yielded greater year-to-year variability in total precipitation. In my opinion, this result should be discussed in the context of existing literature on future climate variability in the Mediterranean area. Which could be the potential role of bias correction in this result? For example, Maraun (2013) investigated the role of bias correction in modifying relative trends in annual precipitation maxima from a RCM and found that the RCM underestimated observed variability, which led to substantial amplification by quantile mapping of modeled trends in extremes. Besides, it would be interesting to examine the future trends obtained from the rest of the GCM/RCM combinations.

Maraun, D. (2013). Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. J. Climate, 26, 2137–2143, doi:10.1175/JCLI-D-12-00821.1

Pendergrass, A.G., Knutti, R., Lehner, F. et al. Precipitation variability increases in a warmer climate. Sci Rep 7, 17966 (2017). https://doi.org/10.1038/s41598-017-17966-y

Polade, S. D., Gershunov, A., Cayan, D. R., Dettinger, M. D., & Pierce, D. W. (2017). Precipitation in a warming world: Assessing projected hydro-climate changes in California and other Mediterranean climate regions. Scientific reports, 7(1), 10783. https://doi.org/10.1038/s41598-017-11285-y

Thank you to the Reviewer comments we realized that this sentence was wrong. Both RCMs forecast a decrease of the standard deviation of precipitation. It is not equivalent to the variability. We have calculated the coefficient of variation of the historical and future series obtained with RCM2 and RCM9 and we obtained 0.80, 1.07, and 1.10 respectively. Therefore, the variability is higher in the future. We have modified this sentence and included the references suggested by the Reviewer.

The considered RCMs predict significant reductions of mean precipitation (-31.6 % and -44.0 % for RCM2 and RCM9 respectively) and increase of mean temperature (26.0 % and 32.2 % for RCM2 and RCM9 respectively) (see Fig. 10a and 10b respectively). The average change in monthly standard deviation of precipitation is -6.2 % and -32.3 % for RCM2 and RCM9 respectively. In the case of temperature these changes are 23.9 % and 4.8 %. Both RCMs predicts a decrease of the standard deviation in precipitation and an increase of the standard deviation of temperature in the future (see Fig. 10c and 10d respectively). However the expected values of changes are significantly different. Both RCMs also predict significantly different changes in the skew coefficient of series (Fig. 10e and 10f). With respect the hydrology analysis, both RCMs predict significant decreases of mean streamflow (-43.5 % and -57.2 % for RCM2 and RCM9 respectively) (Fig. 11a). In the case of the standard deviation, the RCMs predict a reduction (Fig. 11b). The average change in monthly standard deviation is -26.2 % and -57.5 % for RCM2 and RCM9 respectively. In the case of the skew coefficient both RCMs show an increment with respect the historical scenario (Fig. 11c). We also analysed the coefficient of variation (ratio of the standard deviation to the mean) of historical and future series of precipitation, temperature, and streamflow (Table 4). Both RCMs predict an increase of the precipitation and streamflow variability, and a reduction of temperature variability. This increment in precipitation variability is also described in other climate change impact studies (Pendergrass et al., 2017; Polade et al., 2017).

	Coefficient of variation (CV)		
	Precipitation	Temperature	Streamflow
Historical	0.80	0.46	0.69
RCM2	1.07	0.41	0.84
RCM9	1.10	0.42	1.07

Table 4: Coeficient of variation of the historical and future series generated from RCM2 and RCM9 for the precipitation, temperature, and streamflow.

Line 211: "predict significant decreases of streamflow (-43.5 and 57.2%)" Should it be - 57.2%?

Thank you. We corrected it:

With respect the hydrology analysis, both RCMs predict significant decreases of mean streamflow (-43.5 % and -57.2 % for RCM2 and RCM9 respectively) (Fig. 11a).

Lines 215-217: "In the case of the meteorological droughts the first SPI threshold for which droughts periods are detected in the historical scenario is -3.0. In the future scenarios this value is -5.2 and -4.6 for the RCM2 and RCM9 respectively". I think that it will be interesting to assess the changes in the parameters of the future distribution with regard to the historical one (even if only the historical distribution is used to obtain the future SPI).

Done:

Significant changes are also expected for droughts. In the case of the meteorological droughts the first SPI threshold for which droughts periods are detected in the historical scenario is -3.0. In the future scenarios this value is -5.2 and -4.6 for the RCM2 and RCM9 respectively (Fig. 12). In order to perform an appropriate analyses of the future droughts with respect to the historical, the future SPI calculation were estimated by using the parameters of the gamma distribution obtained in the historical period (Collados-Lara et al., 2018). If the parameters of the gamma distribution were adjusted to the future series of values, the changes in the parameters would be significant. For RCM2 we would obtained $\alpha = 19.9$ and $\beta = 2.6$ (instead of the historical values $\alpha = 16.1$ and $\beta = 3.2$).

Lines 219-224: Check the signs of the SPI values.

Done. When we refer to thresholds of SPI we used sing (-) and for statistics (intensity or magnitude) we used (+).

Lines 231-257: in my opinion, the Discussion section does not address properly the limitations of the selected methodology (see my previous comments).

We included a new section in Discussion to point the limitations and future works related to the proposed approach:

5.1 Hypothesis assumed, limitations and future works

Although we have demonstrated the utility of the proposed approach to assess future impacts on meteorological and hydrological droughts, we want to highlight some hypothesis and limitations assumed and to identify potential future research aligned with this study:

- We have used a bias correction method based on the assumption of bias stationarity of climate model outputs. However, this assumption may not be valid to study some problems due to the significance of the influence of climate variability on them. Other approaches should be explored to take into account non-stationarity bias of RCMs simulations (e.g. Hui et al., 2020).
- We have applied the same bias correction procedure for all the range values in accordance with the climatic variable distribution function. We did not consider

the impact of bias correction techniques on the tails of the distribution, which could be important to analyse extremes (Volosciuk et al., 2017).

- In this work a univariate bias correction method is used. It does not consider the dependence between precipitation and temperature which could be explored in future assessments. Meyer et al. (2019) found that incorporating or ignoring inter-variable relationships between temperature and precipitation could impact the conclusions drawn in hydrological climate change impact studies in alpine catchments.
- The streamflow information available for this case study cannot be divided into two long-enough (e.g. 30 years) series representative of the climate/hydrology to perform explicitly a validation of the bias correction models (Chen et al., 2021). We assumed that the statistics of any long-enough periods remain invariant. In this case the calibration implicitly could be considered validated, due to the fact that the same results would be obtained under this hypothesis for any other period representatives of the climate/hydrology conditions.
- In our case study the influence of temperature was considered only in the hydrological assessment by using rainfall-runoff models. However other meteorological droughts indices that consider temperature could be included in the analysis [e.g. the Standardised Precipitation-Evapotranspiration Index (SPEI) (García-Valdecasas Ojeda et al., 2021)].
- 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.
- The proposed method has not been tested in other typologies of basin, as for example in Alpine basins where snowmelt component may have a significant influence on the results.