Response to RC2:

Thank you for inviting me for reading this article. The authors evaluate terrestrial water storage under SSP5-8.5 and SSP5-8.5-SAI scenarios across the Middle East and North Africa. The results are useful for supporting aerosol intervention strategy against global warming and water resources management for Mediterranean, Middle East, and North Africa. I have some concerns about the methods and figures which may be helpful for improvement.

Reply: We sincerely appreciate your effort and time in reviewing our manuscript as well as your constructive comments/suggestions. We have made every effort to incorporate your feedback effectively. Below, you will find a detailed response to each comment, with comments presented in black and our responses in red.

1- Section 2.3. The authors calculate return periods from GEV distribution. However, GEV distribution is used to simulate maximum value in a certain period, instead of monthly values. The authors may give more details about how to apply GEV distribution. Did the authors calculate the annual maximum TWS values? In addition, authors may provide empirical probabilities and examine whether annual maximum TWS follows GEV distribution or other distributions.

Reply: We applied a GEV distribution to the complete dataset of monthly TWS values without explicitly setting maximum values. This approach allowed us to estimate the parameters of the GEV distribution using the entire dataset. However, in response to your request, we have also extracted the annual maximum TWS values and provided the corresponding fitted GEV distribution for comparison with the full dataset scenario (e.g., Figures RC2-1 and RC2-2 below).

Overall, the probability densities for both datasets exhibit a high degree of similarity across various regions and scenarios. For instance, Figures RC2-1 and RC2-2 illustrate the probability densities for the R2 and R5 regions. Additionally, the graphs depicting return levels versus return periods based on annual maximums (Figure RC2-3) closely resemble the results obtained from the entire dataset (Figure 5). In all cases, the trends are highly similar (compare Figure 5 and Figure RC-3), although it's worth noting that the annual maximums scenario exhibits slightly wider upper and lower bounds compared to the entire dataset scenario. Regarding the significance test for differences between historical, global warming, and SAI scenarios, the results are consistent across all cases. However, there is an exception in the case of the difference between historical and SAI scenarios in R5. In the entire dataset scenario, a significant difference is observed (Fig. 5e), whereas in the annual maximums case, it does not reach significance (Fig. S9e).

In light of these explanations, we have retained the results obtained from the entire dataset in the main text, and we will include the results from the annual maximums scenario in the Supplementary Materials.

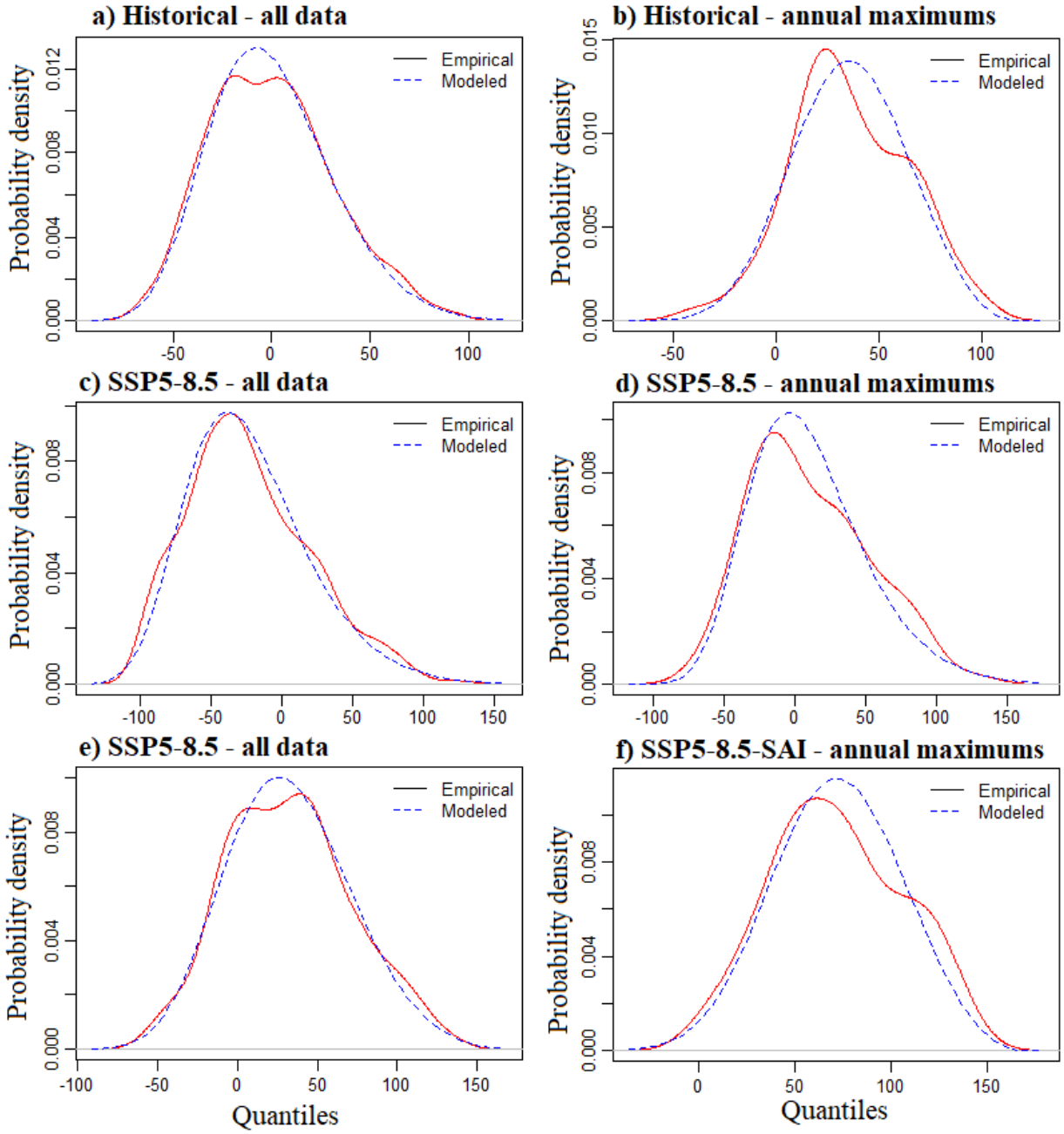


Figure RC2-1. Probability density curves for Region R2, comparing two scenarios: one using all available data (left column) and the other using annual maximum values (right column) under the historical conditions (upper row) as well as the GHG emissions (middle row) and SAI (lower row) scenarios for region R2.

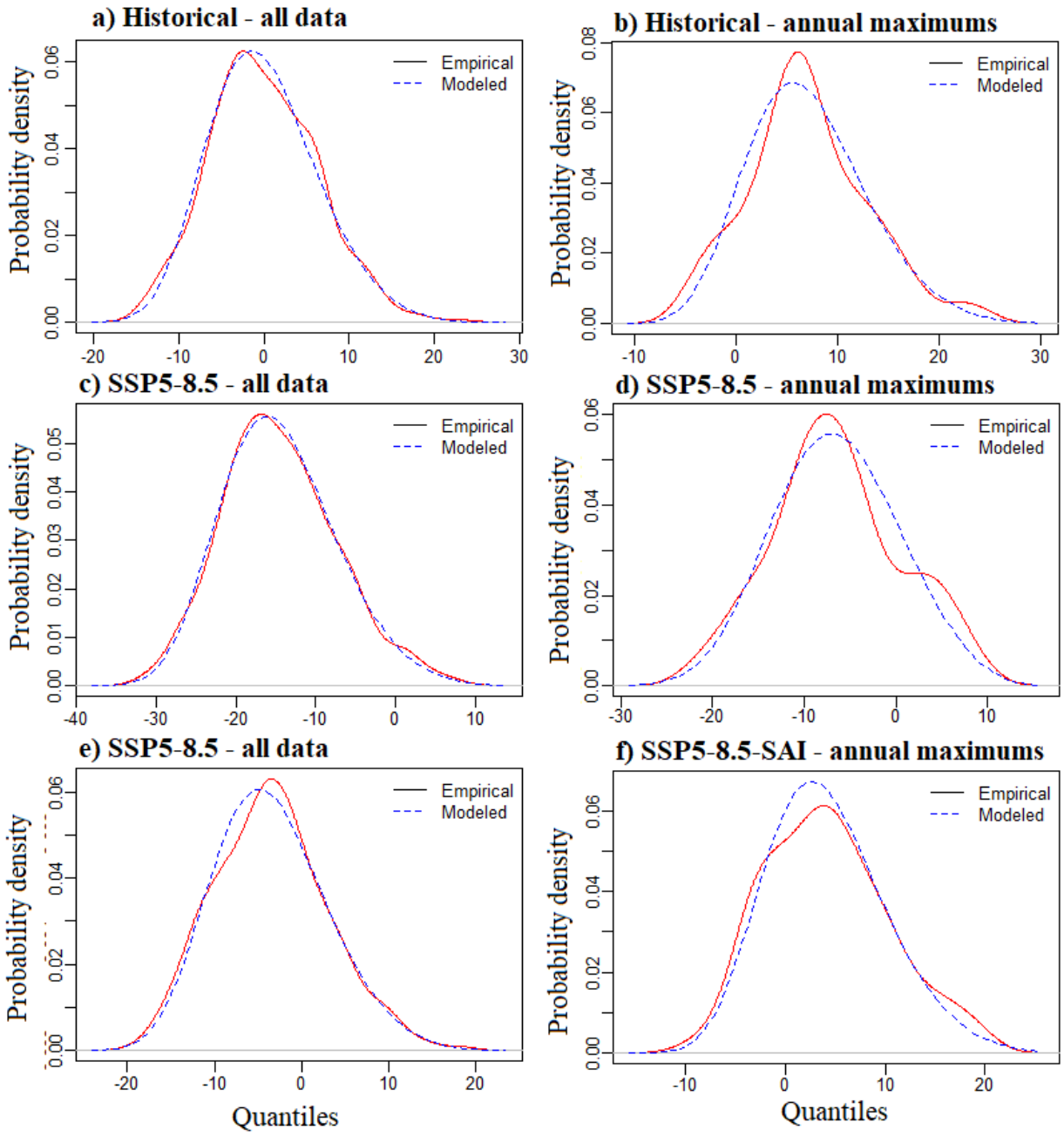


Figure RC2-1. As Figure RC2-1 but for the region R5.

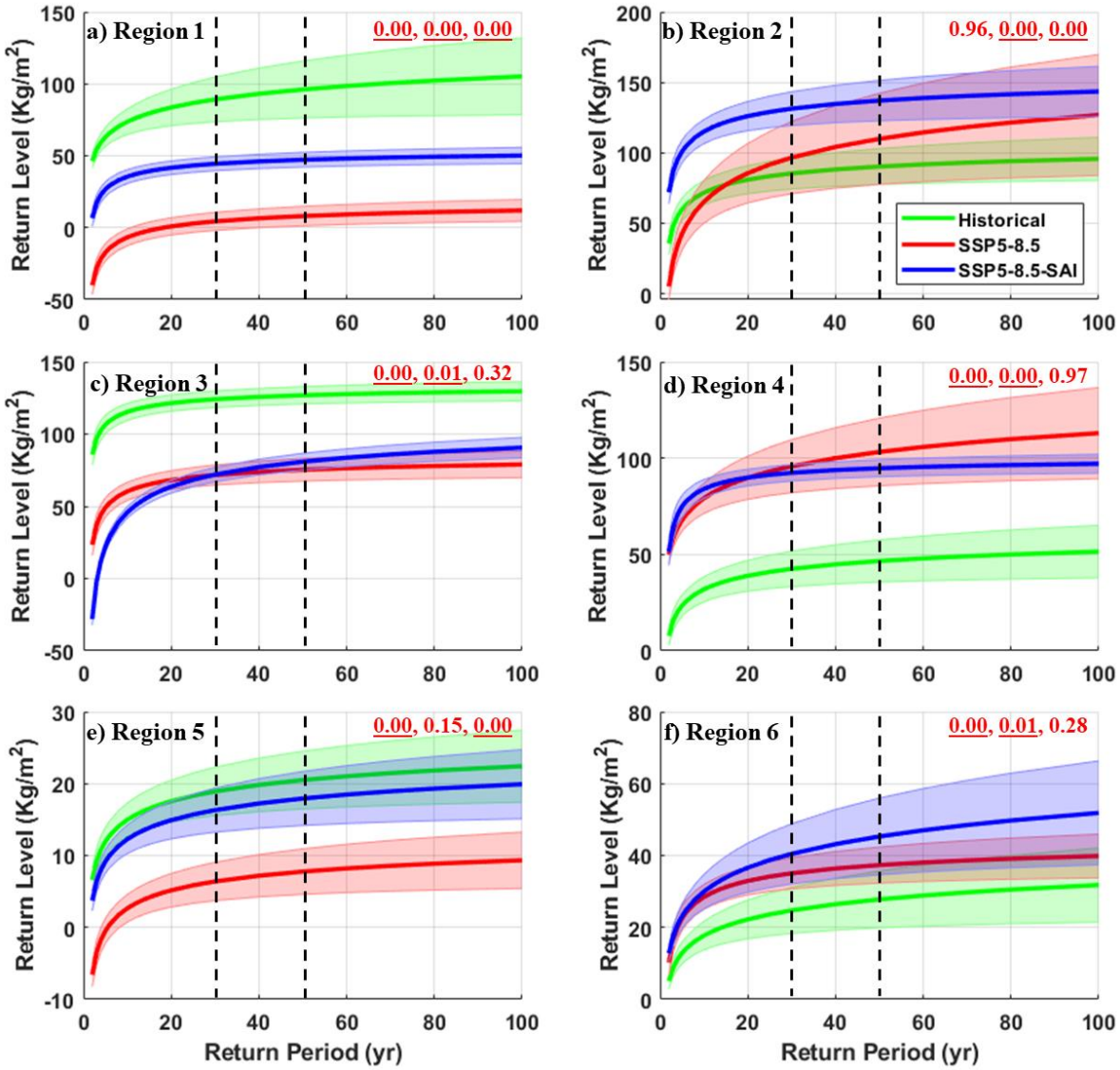


Figure RC2-3. As in Figure 5 but for the annual maximums.

2- Line 211. The historical period is from 1985-2014, and future period is from 2071-2100. The authors do not analyze mid-21th century. The authors may explain why you do not analyze the full period from 1850-2100.

Reply: Agreed. We have included the following clarifications in the text:

"We have chosen to focus on the 2071-2100 future period because the anticipated changes in TWS driven by GHG emissions are expected to be more pronounced during this time frame (Pokhrel et al., 2021). Our decision to prioritize this period is based on the need to examine the effects of more significant alterations resulting from global warming."

"We focused on the historical period from 1985 to 2014 rather than the entire historical dataset spanning from 1850 to 2100 for several reasons. Firstly, recent historical climate data may exhibit less uncertainty, given that additional meteorological stations with improved data quality are available to be used for model calibrations (Zhang et al., 2020). Secondly, this selected historical period offers valuable insights into the observable impacts of climate change, which are highly pertinent to present-day societal and environmental challenges. These insights are of utmost importance to policymakers and communities alike. Thirdly, the chosen historical 30-year time period aligns with the 30-year periods considered for the GHG emissions and SAI scenarios, ensuring consistency in our statistical analysis."

Ref:

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., ... & Wada, Y. (2021). Global terrestrial water storage and drought severity under climate change. Nature Climate Change, 11(3), 226-233.

Zhang, B., Xia, Y., Long, B., Hobbins, M., Zhao, X., Hain, C., ... & Anderson, M. C. (2020). Evaluation and comparison of multiple evapotranspiration data models over the contiguous United States: Implications for the next phase of NLDAS (NLDAS-Testbed) development. Agricultural and Forest Meteorology, 280, 107810.

3- Authors only select CESM2 for analysis. The authors may evaluate the performance CESM2 for historical climate over the study area to validate this model.

Reply: We have incorporated the following information into the new version:

"For global terrestrial ET, the CESM2(WACCM) ranked as the second-best model among 19 CMIP6 models (Wang et al., 2021)."

"In the evaluation by Babaousmail et al. (2021), which assessed 15 CMIP6 models in replicating monthly rainfall patterns spanning from 1951 to 2014 in NA, CESM2(WACCM) emerged as one of the top-performing models. It accurately captured rainfall peaks across the region, albeit with a slight overestimation (ranging from 5 to 10 mm/month) in the southern areas and a slight underestimation (ranging from 0 to 20 mm/month) in the northern regions. Despite these minor deviations, CESM2(WACCM) was recognized as one of the better models for simulating precipitation patterns across North America, achieving a Taylor skill score (TSS) of 0.62. Evaluation of CESM2(WACCM) across the Mediterranean coasts placed it at the 9th and 17th positions out of 31 CMIP6 models for its performance in simulating temperature and precipitation (Bağçaci et al. (2021). Furthermore, when it comes to simulating precipitation relative to observational data for northeastern Iran during the period of 1987-2005, CESM2 stood out as the top-performing model among six CMIP6 models (Zamani et al., 2020). Assessing the representation of spatial and temporal variations in historical

precipitation from 1980 to 2014 across Africa and the Arabian Peninsula, the CMIP6 multi-mean ensemble (inclusive of CESM2-WACCM) demonstrated reasonable performance, as highlighted in Nooni et al. (2023)."

Ref:

Babaousmail, H., Hou, R., Ayugi, B., Ojara, M., Ngoma, H., Karim, R., ... & Ongoma, V. (2021). Evaluation of the performance of CMIP6 models in reproducing rainfall patterns over North Africa. Atmosphere, 12(4), 475.

Bağçaci, S. Ç., Yucel, I., Duzenli, E., & Yilmaz, M. T. (2021). Intercomparison of the expected change in the temperature and the precipitation retrieved from CMIP6 and CMIP5 climate projections: A Mediterranean hot spot case, Turkey. Atmospheric Research, 256, 105576.

Nooni, I.K.; Ogou, F.K.; Chaibou, A.A.S.; Nakoty, F.M.; Gnitou, G.T.; Lu, J. (2023). Evaluating CMIP6 Historical Mean Precipitation over Africa and the Arabian Peninsula against Satellite-Based Observation. *Atmosphere*, *14*, 607. https://doi.org/10.3390/atmos14030607

Wang, Z., Zhan, C., Ning, L., & Guo, H. (2021). Evaluation of global terrestrial evapotranspiration in CMIP6 models. Theoretical and Applied Climatology, 143, 521-531.

Zamani, Y., Hashemi Monfared, S. A., Azhdari Moghaddam, M., & Hamidianpour, M. (2020). A comparison of CMIP6 and CMIP5 projections for precipitation to observational data: the case of Northeastern Iran. Theoretical and Applied Climatology, 142, 1613-1623.

4- Authors use the MLR model to predict TWS. Apart from potential ET, the actual ET is also correlated with temperature and precipitation. How to solve the collinearity between ET, temperature and precipitation?

Reply: In assessing collinearity, we employed the VARCLUS procedure, a method that partitions a set of numeric variables into distinct or hierarchical clusters (Sarle, 1990). Each cluster is associated with a linear combination of the variables it contains. The criterion in this procedure is that if the proportion of the variance explained by a cluster is larger than 0.8 (Figures RC2-4 and RC2-5 below), we should choose one variable from that cluster. It's worth noting that there was minimal variation among ensemble members for each scenario across regions. As a result, we have exclusively presented results for the ensemble r1 in Figures RC2-4 and RC2-5. Based on our findings (refer to Figures RC2-4 and RC2-5 below), in most instances, we needed to select one variable from pairs like potential ET and temperature or TWS and soil moisture. Consequently, we opted for temperature and TWS for our analysis. However, in the arid regions R4 to R6, although both real ET and precipitation were categorized within a single cluster, the proportion of variance explained by the cluster fell below 0.8. Hence, we decided to consider both variables. In response to the comment made in RC1, we also included leaf area index (LAI) as an additional variable in our analysis.

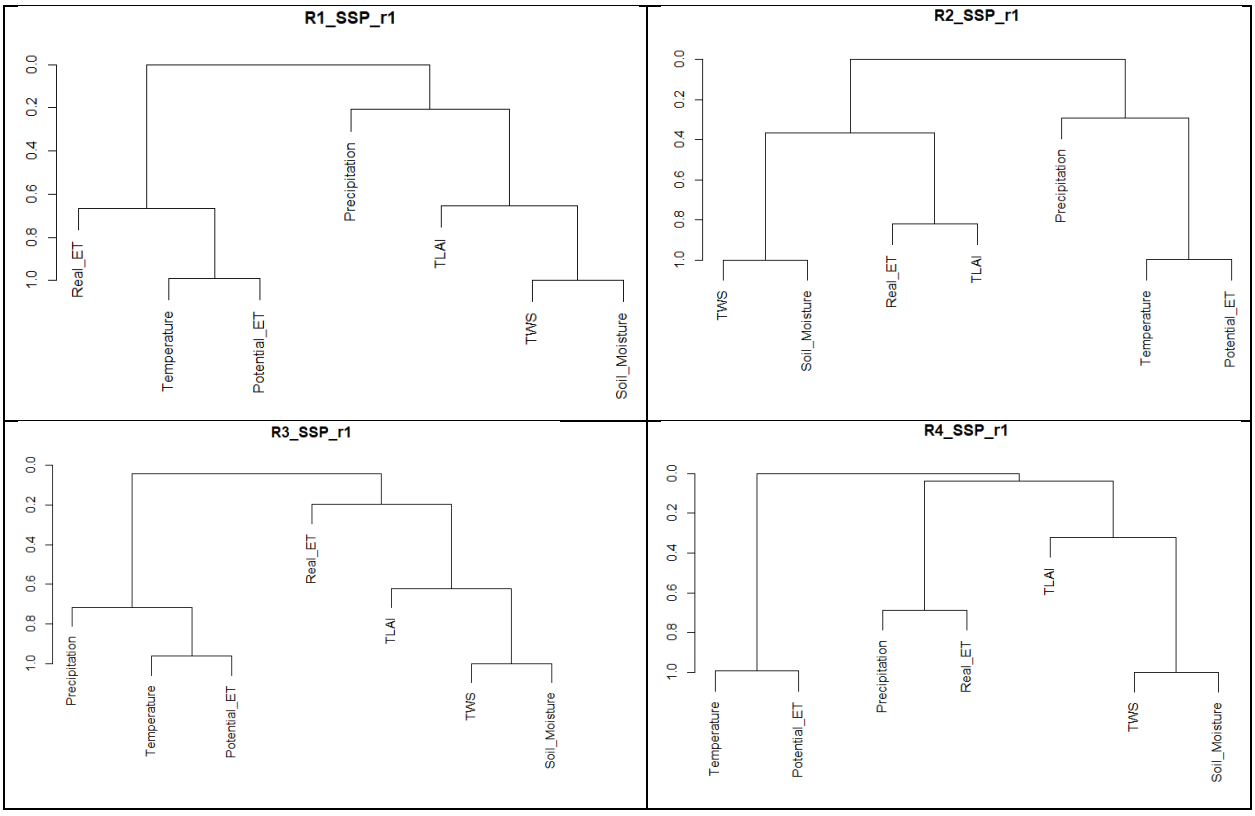
Incorporating above additional explanations into Section 2.4 of the methodology, it would read as follows:

"We employed the VARCLUS procedure to thoroughly assess collinearity among the variables. VARCLUS is a method that effectively segregates a set of numeric variables into disjoint or hierarchical clusters, each characterized by a linear combination of the variables within the cluster (Sarle, 1990). The criterion is that when the proportion of the variance explained by a cluster is larger than 0.8, it is advisable to select one variable from that cluster.

Based on the results obtained from VARCLUS, we made specific decisions to enhance the robustness of our analysis. For instance, we identified strong correlations exceeding 0.9 between potential ET and temperature, as well as between soil moisture and TWS. Consequently, we chose to exclude potential ET and soil moisture from our analysis due to their high levels of correlation with temperature and TWS, respectively."

Ref:

Sarle, W. (1990). The VARCLUS Procedure. In *SAS/STAT User's Guide* (fourth, Vol. 2, pp. 1641–1659). SAS Institute, Inc.
http://support.sas.com/documentation/onlinedoc/stathttp://support.sas.com/documentation/onlinedoc/stat



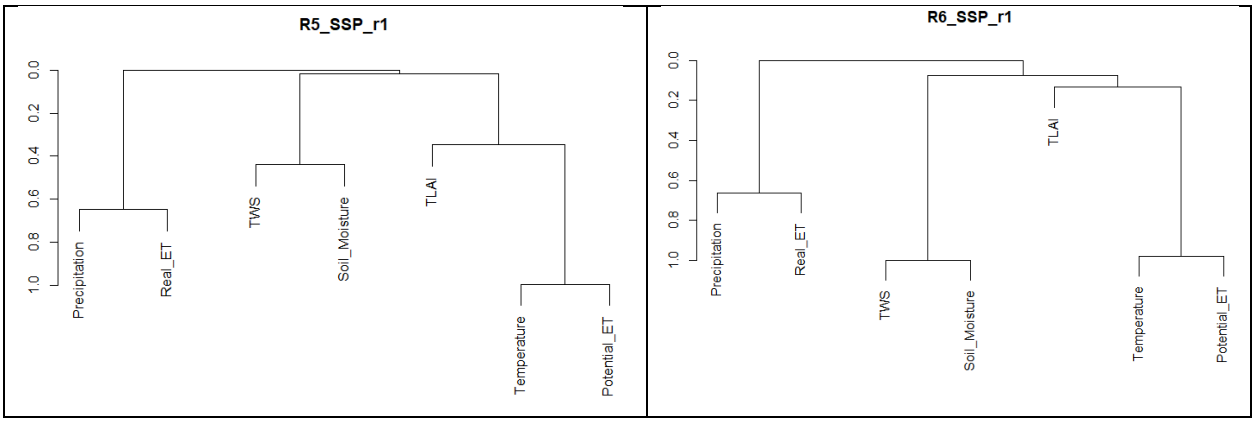
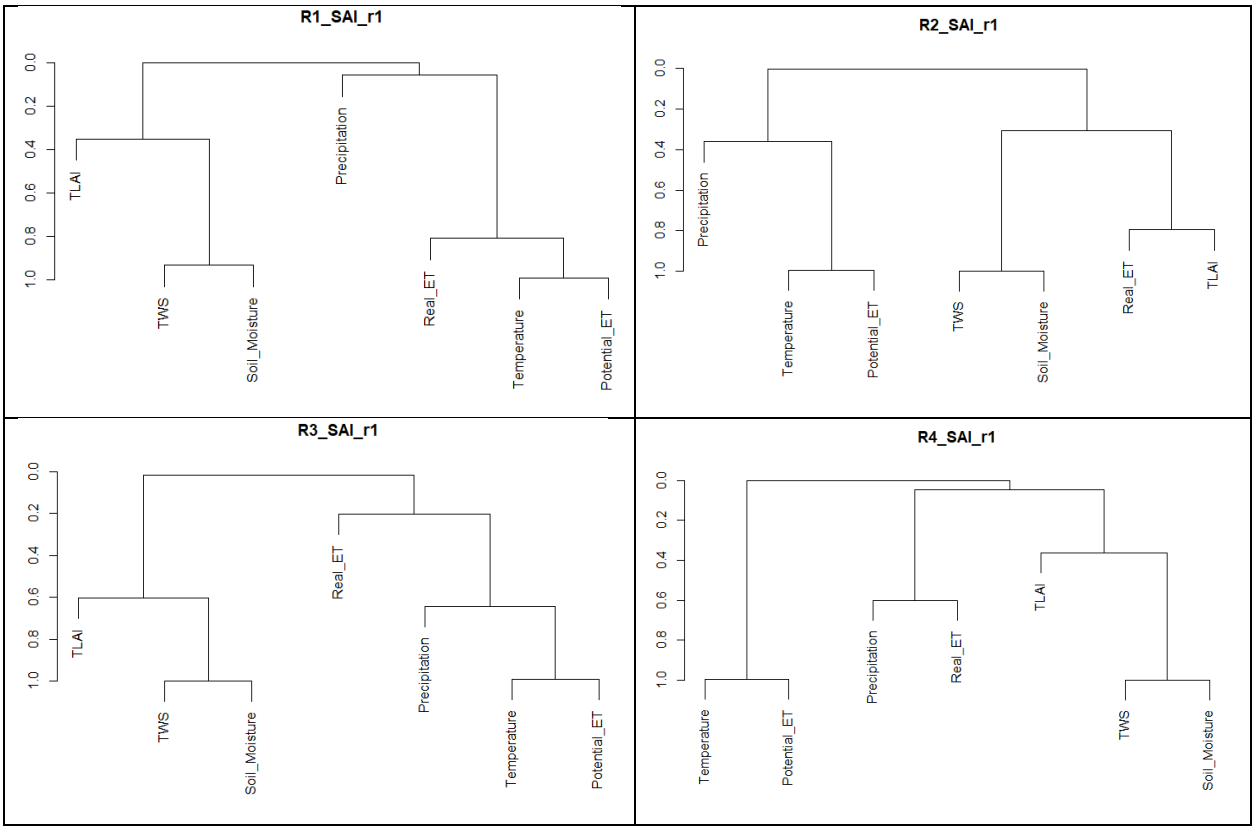


Figure RC2-4. This tree diagrams illustrate the cluster hierarchy within ensemble r1 of the SSP5-8.5 scenario across regions R1 to R6. The y-axis represents the Proportion of Variance Explained.



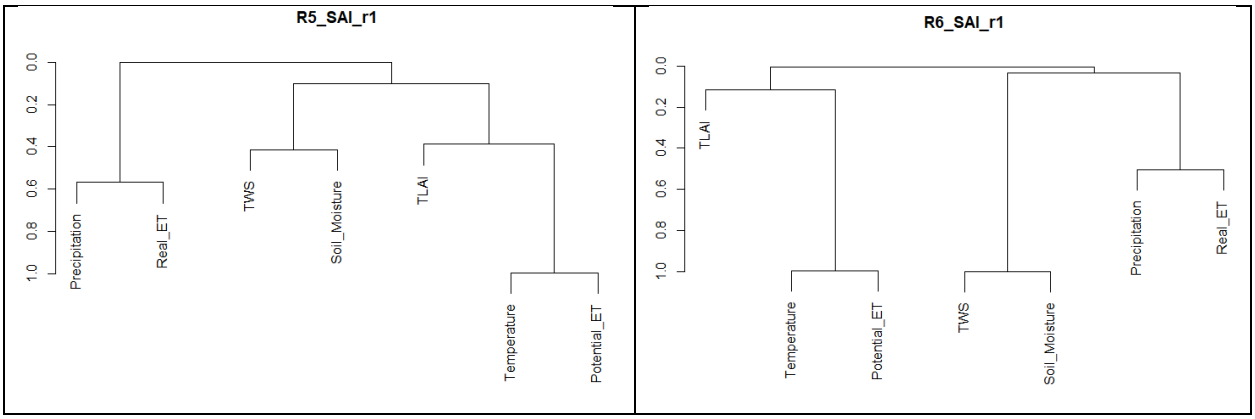


Figure RC2-5. As in Figure RC2-4 but for the SSP5-8.5-SAI scenario.

5- Authors remove outliers in the MLR model. This will artificially give better results. Please justify the removal of these values?

Reply: Overall, the maximum number of outliers removed (5) is relatively insignificant when considering the total number of records in each timeseries (which exceeds 700 in our study). Therefore, it is unlikely to have a substantial impact on the model. Nonetheless, we have incorporated the following statement into the text:

"The number of outliers data points excluded varies from zero to 5 (of the 700 points) in the 36 models"

6- The temporal autocorrelation is an important component in TWS evolution. Monthly TWS is not only impacted by concomitant precipitation and temperature, but also antecedent soil moisture and climatic variables. Authors may consider include climatic variable in previous months as predictors as well.

Reply: In our models, we excluded soil moisture from the list of predictor variables due to its collinearity with TWS. Additionally, we conducted a temporal autocorrelation analysis on all the variables, including temperature, precipitation, real ET, and LAI data for each model. This analysis was carried out using the Autocorrelation function at a 95% confidence level.

In all regions (except R4), the autocorrelation results indicated that the lags at the first and second months were statistically significant, while the third month lag was almost non-significant. Therefore, we modified the LMS model to include information from the two preceding months in these regions.

However, in region R4, we observed different patterns. In this region, both real ET and temperature significantly depended on their respective conditions from the two previous months, while precipitation did not show this effect. Moreover, TLAI in R4 exhibited dependencies on the first three and four preceding months under the SSP and SAI scenarios, respectively. Consequently, we incorporated specific lagged months for each variable in R4.

We have included the updated figures (Figures 6 and 7) below to reflect these changes. Furthermore, we will revise the MLR methodology and Section 3.3 in accordance with this information.

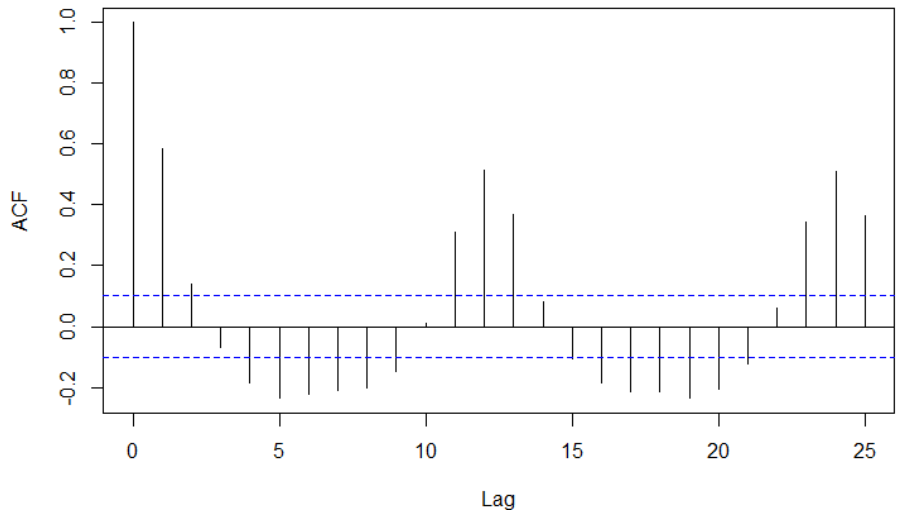


Figure RC2-6. The autocorrelation plot for real ET in region R4 under the SAI scenario, specifically ensemble member 003. The y-axis represents lag values in terms of months.

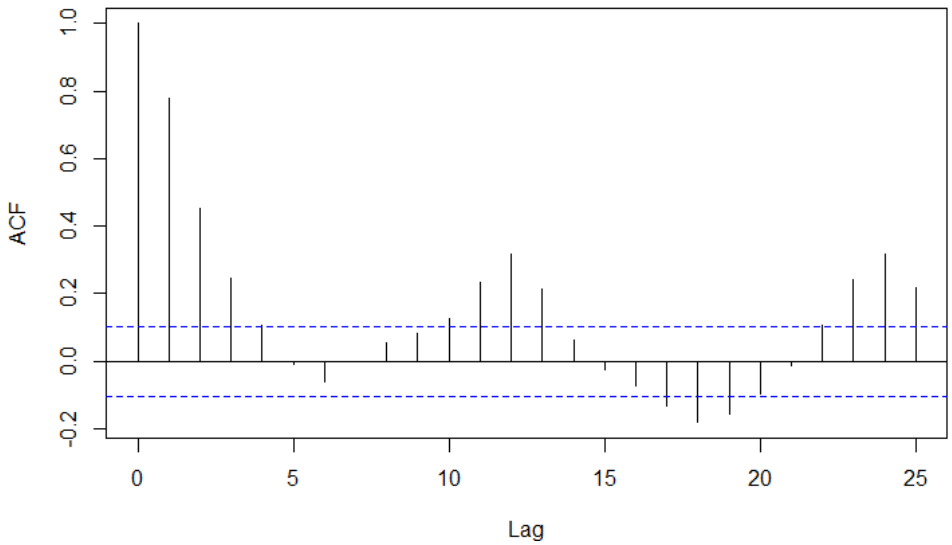


Figure RC2-7. As in Figure RC2-6 but for TLAI.

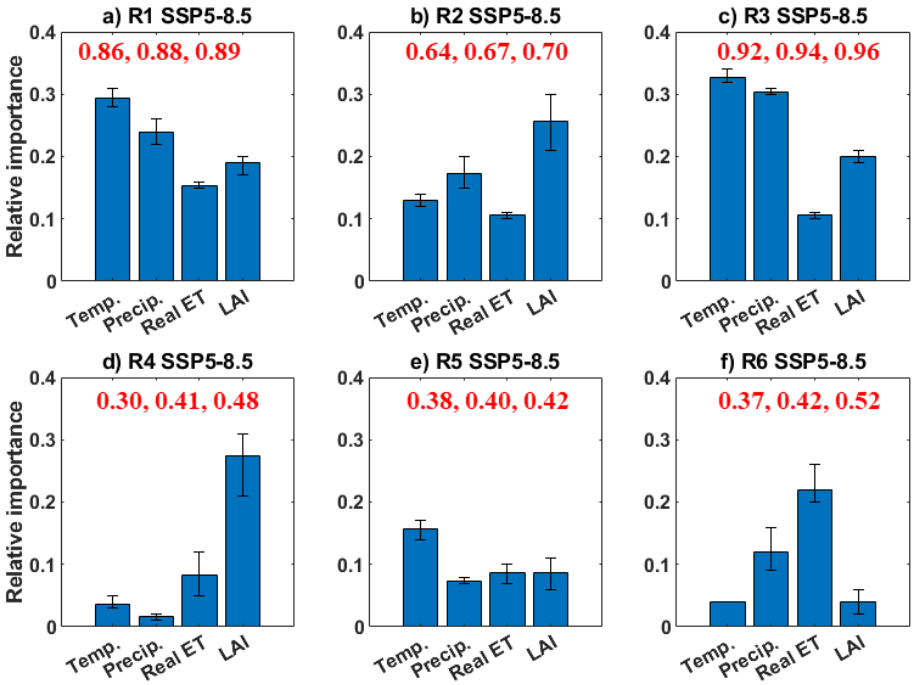


Figure 6. LMG importance plot (Lindeman et al., 1980) of the four independent variables in the regression for TWS for the global warming SSP5-8.5 scenario in each region. The bar and range-bar respectively show the ensemble mean importance and the range of importance from the three ensemble members. The three values in red on each subplot shows the minimum, mean, and maximum variances explained by models.

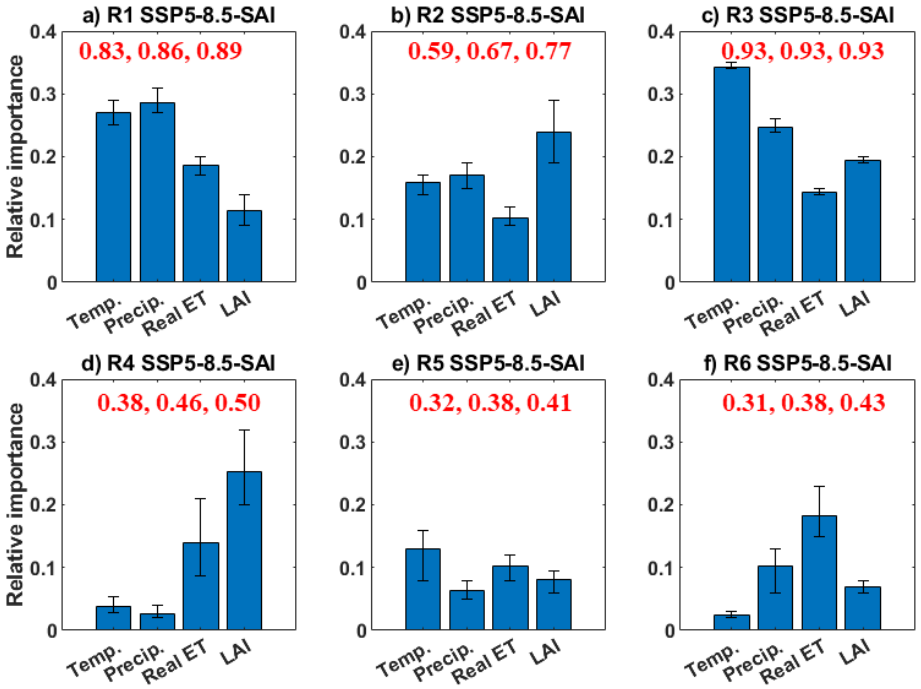


Figure 7. As in Fig. 6, but for the SSP5-8.5-SAI scenario.

7- Water storage include soil moisture, groundwater, snow, ice, and others. Figure S3 seems to indicate soil moisture is the dominant driver of TWS variations. It may be insightful for evaluate the relative contributions of other components in TWS.

Reply: In the CMIP6 climate models, TWS is defined as the sum of snow water equivalent and soil moisture (Wu et al., 2021). Consequently, it is reasonable to assume that soil moisture plays a dominant role in driving TWS variations, especially in arid regions. We have incorporated this information into the text with the following sentence:

"TWS is the sum of snow water equivalent and soil moisture (Wu et al., 2021). In the drier regions the soil moisture variability accounts for the dominant component of TWS variability (Pokhrel et al., 2021)"

Ref:

Wu, R. J., Lo, M. H., & Scanlon, B. R. (2021). The annual cycle of terrestrial water storage anomalies in CMIP6 models evaluated against GRACE data. Journal of Climate, 34(20), 8205-8217.

8- Figure S4 is important for interpreting current results. May consider to place this figure in main text.

Reply: Understood. Since the primary focus of the study is on Terrestrial Water Storage (TWS), it will be kept in the supplementary materials.

9- It may be useful to compare the results with previous evaluations (https://www.nature.com/articles/s41558-020-00972-w; Global terrestrial water storage and drought severity under climate change).

Reply: Implemented. We used it in our discussions as follows:

"The CMIP5 outputs also confirm that the global warming (RCP2.6 and RCP6.0) substantially decreases the TWS in the Mediterranean by the mid- (2030-2059) and late- (2070-2099) twenty-first century (Pokhrel et al, 2021)."

"Similarly, a decrease in precipitation (Kim and Byun, 2009), surface runoff (Cook et al., 2020), and TWS (Pokhrel et al., 2021) has been reported across Mediterranean coasts under GHG warming."

Ref:

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., ... & Wada, Y. (2021). Global terrestrial water storage and drought severity under climate change. Nature Climate Change, 11(3), 226-233.

10- Line 29, this sentence may be improved. May explain "more continental" and "hyper-arid" climates? Specify what is different response?

Reply: Implemented. We edited the text as follows:

"… with hyper-arid climate (with annual precipitation less than 100 mm) has the lowest precipitation, real ET, soil moisture, and TWS. More continental areas have characteristics that are typical of continental climates and are less influenced by the moderating effects of nearby oceans."

To clarify the distinctive response observed in R5, we included the following information in the text: "Unlike the other arid regions, in eastern NA (R5), we observe a reduction in the mean TWS trend under both GHG and SAI scenarios, and the extreme TWS values are also lower compared to the historical conditions."

11- Line 86, may place this paragraph earlier than the introduction of SRM, which is proposed to address climate change.

Reply: Implemented.

12- Line 127. What is the regional consequence and hydrological cycle? May give more explanations

Reply: We revised it as follows:

"While SAI may counteract the annual-mean water availability changes over land forced by GHG, it is not easy to offset the regional consequences, especially in the hydrological cycle, such as the Amazonian drying trend and the reduced precipitation (P), evaporation (E), and P-E (Jones et al., 2018)."

13- Line 228 and Eq. (1). The authors give the equations for Xi = 0 in equation (2). It may be better to provide CDF when Xi (ξ)= 0 in Equation (1) as well. In addition, I think Eq.(1) is the CDF instead of PDF. It is better to clearly specify this.

Reply: Implemented. We added mor explanations to this part as follows:

"The GEV probability density and cumulative distribution functions are defined as (Gilleland, 2020):

$$g(z)=\frac{1}{\sigma}t(z)^{1+\xi}e^{-t(z)};\ G(z)=e^{-t(z)};\ t(z)=\begin{cases}\left\{1+\xi\left(\frac{z-\mu}{\sigma}\right)\right\}^{-1/\xi}, & \xi\neq 0\\ e^{-\left(\frac{z-\mu}{\sigma}\right)}, & \xi=0\end{cases} \quad (1)$$

For $\xi \neq 0$, we have $t(z)^{1+\xi} = \left\{1+\xi\left(\frac{z-\mu}{\sigma}\right)\right\}^{-(1+1/\xi)}$ and for $\xi = 0$, the x domain restricted to $\xi\left(\frac{z-\mu}{\sigma}\right) > -1$. The GEV distribution is parameterized using ξ, μ, and σ which are the shape, location, and scale parameters, respectively and analogous to the skewness, mean and standard deviation."

14- Line 272, this sentence may be improved.

Reply: Agreed. We have rewritten it as follows:

"The TWS difference between SAI and global warming in the region R2, particularly over the latter part of the 21st century, is greater than for the rest of the domain."

15- Figure 3. The colors for legend may be improved. For example, use two different hues to represent positive and negative values, and use white to repesent 0.

Reply: Implemented. As an example, please see new Figure 3.

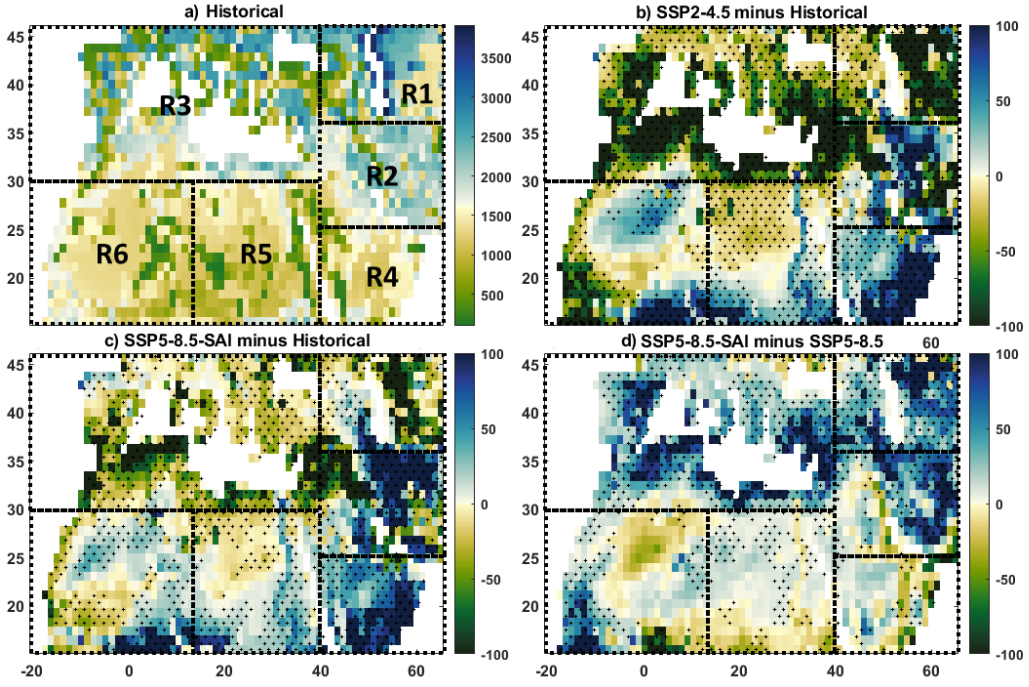


Figure 3. Ensemble mean maps of TWS across the studied domain in the historical climate (a) over 1985-2014 and their projected future changes in the 2071–2100 period under the SSP5-85 GHG scenario (SSP5-8.5 minus historical (b) and GHG+SAI minus historical (c)). The extent to which the SAI impacts the TWS changes imposed by global warming is further shown (SAI minus SSP5-8.5 (d)). Hatched areas show where all ensemble members agree on the sign of the changes.

16- Figure 5, it may be much better to show empirical probabilities of observed TWS and visually show the performance of GEV distribution.

Reply: In response to your request, we have included graphs in the Supplementary Information for the three different scenarios in two regions, R2 and R5, as examples. Due to space constraints, it is not practical to display graphs for all three scenarios in all six regions.

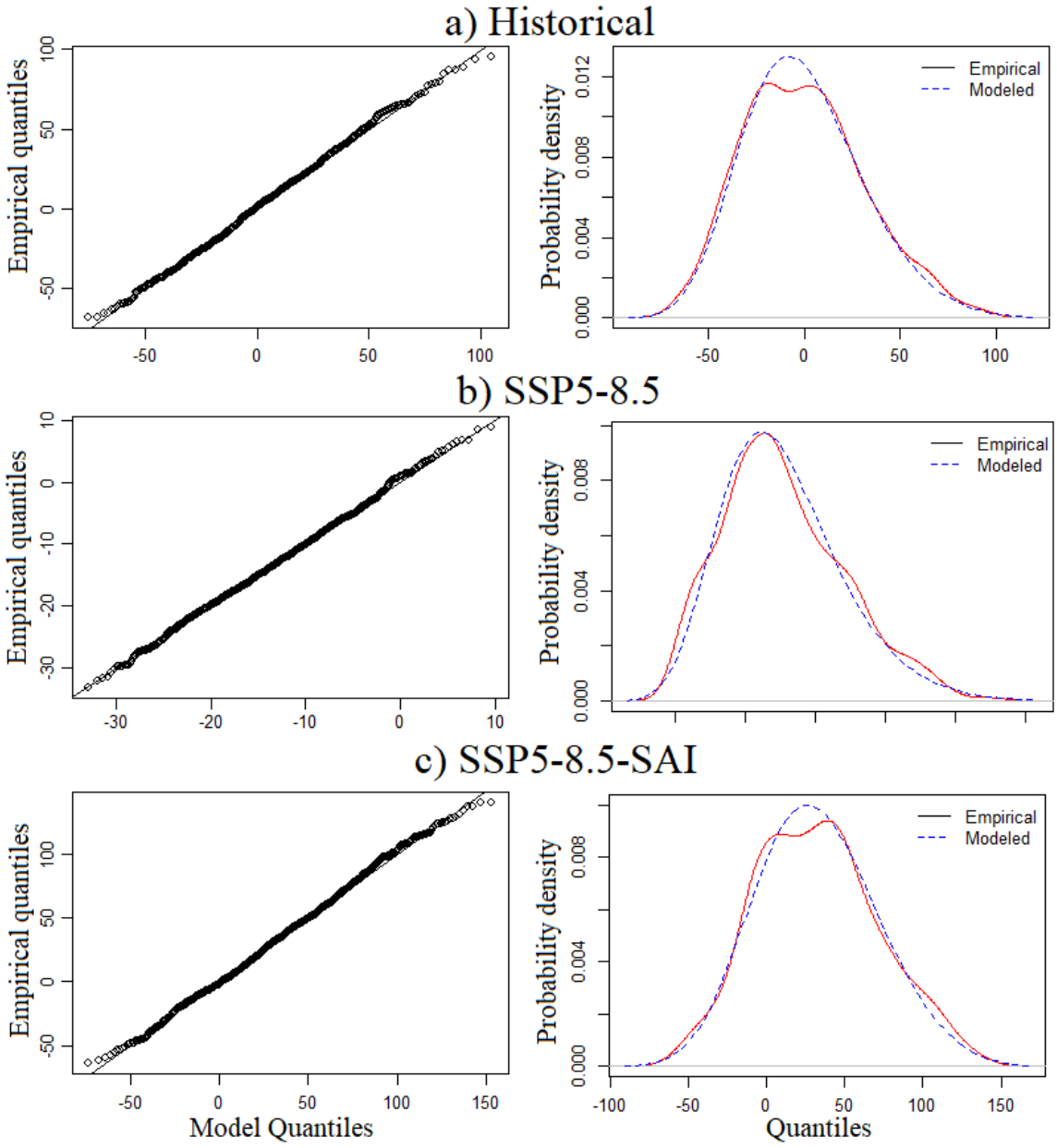


Figure RC2-8. In region R2, the graphs illustrate the following scenarios: (a) historical, (b) global warming, and (c) the SAI scenario. In the left column, you can observe the relationship between empirical quantiles and model quantiles. In the right column, the graphs depict the probability density versus quantiles.

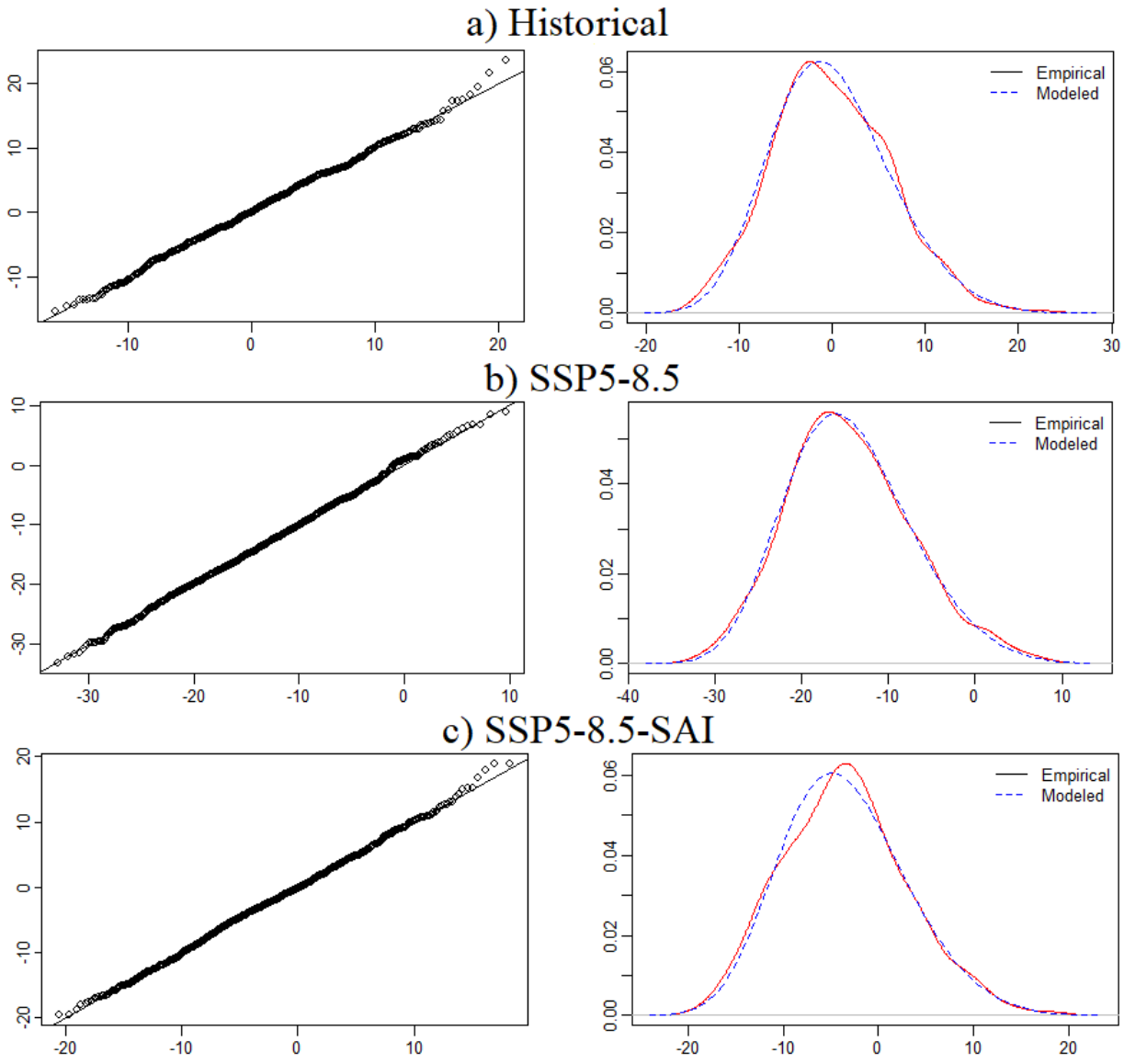


Figure RC2-9. As in Figure RC2-8 but for R5.

17- Figures 6 and 7. May add R-squared in the figures for better interpretation.

Reply: Implemented. Please see the new Figures 6 and 7 above.