A Comparative Analysis of Multi-Model and Downscaled Decadal Climate Predictions over the Southern African Development Community

Author: Sara Moreno Montes[1,2], smorenmo17@alumnes.ub.edu
Supervisors: Carlos Delgado Torres[1] and Albert Soret Miravet[1]
Tutor: Raül Marcos Matamoros[2]
[1]*Barcelona Supercomputing Center (BSC), Barcelona, Spain*
[2]*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain*

Abstract: This study assesses the effectiveness of various downscaling methods applied to multi-model decadal predictions in the Southern African Development Community (SADC). These forecasts combine decadal predictions from the 13 forecast systems contributing to the Decadal Climate Prediction Project Component A (DCPP-A), one of the components of the Coupled Model Intercomparison Project Phase 6 (CMIP6). This work focuses on mean near-surface air temperature and precipitation for the forecast years 1-5. The performance of the different downscaling methods is determined by comparing their forecast quality against raw, coarser-resolution predictions by means of the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) coefficient, which ranks them. It has been found that the ranking primarily depends on the calibration or linear regression approaches, with little differences resulting from the tested interpolation methods. For both variables, the 4 nearest neighbors linear regression method provides the highest skill. However, the outcomes of the additional downscaling process vary between temperature and precipitation. For instance, applying specific downscaling methods for temperature improves skill compared to raw predictions in some areas (i.e. the highest quality is achieved by correcting the mean value and variance, or by correcting just the mean value). For predictions of precipitation, the basic linear regression returns the highest forecast quality, outperforming the rest of techniques, such as simple bias correction and simple interpolation.

I. INTRODUCTION

Until recently, climate projections were the only source of information to estimate climate evolution in the coming years and decades. Nowadays, there is another source of climate information aiming to predict the evolution of the climate system from 1 to 10 years ahead: decadal climate predictions (*Dunstone et al.* 2020). Whereas both climate projections and decadal predictions contain information about external forcings, their main difference is that decadal predictions also include information on the phase of the internal climate variability (*Delgado-Torres et al.* 2022, *Doblas-Reyes et al.* 2013, *Smith et al.* 2019).

In order to incorporate the predictability provided by this internal climate variability, climate models are initialized once per year with observation-based products, known as initial conditions (*Hazeleger et al.* 2013, *Smith et al.* 2013). The goal of this initialization process is to incorporate other sources of predictability that have a slow evolution, such as changes in ocean state or atmospheric composition (*Boer* 2011, *Doblas-Reyes et al.* 2013). Consequently, by incorporating information about the current climate state, the initialization process aligns the phase of the predictions with observations.

This process is also conducted for previous decades, obtaining a set of past predictions known as hindcasts. These hindcasts are essential tools for estimating forecast quality and determining its usability and value for decision-making in climate-dependent sectors such as agriculture (*Solaraju-Murali et al.* 2021), renewable energy (*Bruno Soares et al.* 2018) and water management (*Paxian et al.* 2019). The estimation of the forecast quality is achieved by comparing the hindcasts with observations to determine how well the model has performed in predicting past conditions.

However, climate predictions are also impacted by different kinds of errors as, for example, by inaccuracies in the initial conditions, which arise due to observational uncertainty (*Slingo and Palmer* 2011). To account for these errors, the ensemble approach is used, which involves producing multiple predictions with slight perturbations in the initial conditions, known as ensemble members. This approach provides different paths for the evolution of the climate system and, hence, it gives an estimate on the uncertainty associated with the predictions.

Another source of error is the uncertainty inherent to the models themselves, stemming from the inadequate representation of the climate system, the use of approximate mathematical techniques to solve physical equations, and the systematic errors present in all forecasting systems. For instance, one of this errors is the model's drift, which occurs when the system is initialized with initial conditions that deviate from the model's preferred state, and the forecasts gradually progress towards that state (*Boer et al.* 2016). In that case, to minimize the impact of these model-specific errors, a multi-model approach can be used. This approach is expected to provide higher-quality predictions due to: (i) the cancellation of errors between models, (ii) increased ensemble size, and (iii) the signal that each model adds to the ensemble (*DelSole and Tippett* 2014).

Beyond merely serving as a tool to evaluate model skill, hindcasts play a critical role in identifying these systematic errors, also known as model biases. Recognizing these biases is vital as it enables the application of bias adjustment and calibration techniques, which partially correct these errors and, thus, improve the model's overall accuracy. These adjustments ensure the statistical parameters of the predictions, such as the mean and variance, to be more similar to the observed values, improving their potential applicability (*Doblas-Reyes et al.* 2005, *Dunstone et al.* 2020).

Additionally, since producing decadal predictions is very computationally expensive, models are often run at a lower resolution than what might be optimal for user-specific needs. For instance, a grid point may not entirely represent a specific location if there are features like mountains and valleys within that same grid cell. To overcome this limitation, downscaling methods can be applied to increase the resolution of the predictions and, therefore, provide users with regional information better suited for their local decision-making frameworks.

Currently, two main types of downscaling approaches exist: dynamic and statistical (*Gutierrez et al.* 2013). Dynamical downscaling involves running at higher-resolution climate models to simulate local processes (*Keller et al.* 2022). In contrast, statistical downscaling, which we have used in this work, uses different statistical techniques to increase the resolution of the raw model output, thereby providing regional information of future climate (*Fowler et al.* 2007).

Considering all these aspects, this work seeks to evaluate and compare different post-processing methods to improve the performance and resolution of decadal forecasts for the Southern African Development Community (SADC) within the FOCUS-Africa project (https://focus-africaproject.eu/). This project aims to develop tailored climate services in the SADC region to adapt to the impacts of climate variability and change. Particularly, one of its key objectives is to evaluate the forecast quality, and analyze and compare various calibration and downscaling techniques in order to identify which ones are the most effective. These methods will then be applied to case studies for different climate-vulnerable sectors to ensure the delivery of the highest quality climate services possible for regional decision-making. In addition, up to our best knowledge, there is no study systematically analyzing the performance of different calibration and downscaling methods for decadal predictions over Africa.

II. DATA

In this study, we have utilized decadal predictions from the Decadal Climate Prediction Project Component A (DCPP-A; *Boer et al.* (2016)) of the Coupled Model Intercomparison Project Phase 6 (CMIP6; *Eyring et al.* (2016)), which provides the latest and most comprehensive set of hindcasts resulting from a collaborative effort involving global climate modeling centers. The characteristics and information of these forecast systems is presented in S1. This work employs these hindcasts to analyze and understand the skill of these systems in predicting climate variability at decadal time scale.

In order to evaluate the skill of raw and downscaled predictions, we have used the high-resolution ERA5land reanalysis (*Muñoz Sabater et al.* 2021) as the reference dataset. This reanalysis has been chosen due to its high spatial resolution (0.1°, which corresponds to 11.1 km in the equator).

The variables considered are the mean near-surface air temperature, mean precipitation and three extreme indices based on daily minimum and maximum temperature and precipitation (TN10p, TX90p, R95p). TN10p corresponds to the yearly percentage of days when minimum temperature is below the 10th daily percentile; TX90p, provides the annual percentage of days when maximum temperature is above the 90th daily percentile; and R95p, gives the annual sum of precipitation in days where daily PR exceeds the 95th percentile of daily precipitation (*Zhang et al.* 2011). The evaluation period ranges from 1961 to 2014 for two forecast periods: one year (start dates 1960-2013), and years 1-5 (start dates 1960-2009). Nevertheless, given the length constraints and in the interest of conciseness, we will focus our discussion on the main outcomes for precipitation and temperature for the forecast years 1-5.

III. METHODOLOGY

In this section, we describe the different post-processing techniques that have been applied to the decadal forecasts with the aim to increase their potential usability by the end-users. The first step consisted on setting up a multi-model with all the different decadal models and ensembles. On this regard, we first brought all the models to the same common resolution. Thus, considering that the coarsest model was the CanESM5 with 2.8°, all the models have been interpolated to that resolution before assembling the multi-model. Afterwards, we have built the multi-model using the multi-model-mean approach, assigning equal weight to every model regardless of its number of members. This choice was based on the similarity of results obtained with other multi-model approaches (*Delgado-Torres et al.* 2022).

After that, we have applied different downscaling approaches to the multi-model in order to bring the predictions to the ERA5land resolution (0.1°). All these computations have been done in cross-validation, that is, excluding data from the specific year under post-processing to avoid overfitting and, consequently, the overestimation of the actual skill. The downscaling methods applied are the following:

1. Interpolation: this approach is widely employed to obtain higher resolution grids by estimating their values from the coarser grid cells. We have selected

five interpolation methods: two conservative variants (first and second order), bilinear, bicubic, and nearest neighbor. Since spatial interpolation can be understood as a kind of zero-order downscaling, it has been used as a benchmark for evaluating more advanced downscaling approaches (*Schoof* 2013).

2. Interpolation plus bias adjustment: this method involves combining the initial interpolation with different calibration techniques: simple bias adjustment, which corrects the mean value (*Leung et al.* 1999); bias adjustment (EVMOS; (*Van Schaeybroeck and Vannitsem* 2011)), that rectifies the mean value and the variance; Mean Square Error minimization (MSE-min; (*Doblas-Reyes et al.* 2005)), that corrects the mean value, the variance and the spread (using two parameters); Continuous Ranked Probability Skill minimization (CRPS; *Gneiting and Raftery* (2005)), that uses three parameters to do the correction; and ratio predictable components (RPC-based calibration; *Eade et al.* (2014)), which aims to increase the signal-to-noise ratio. In addition, quantile mapping has been also applied for precipitation. This technique adjusts simulated data by matching the percentiles of observed and simulated distributions (*Fang et al.* 2015).

3. Interpolation plus linear regression: this approach integrates both interpolation and linear regression in the downscaling process. Two distinct methodologies have been compared for the linear regression step. The first method utilizes high-resolution observations as predictands and the interpolated model as the predictor (for the same variable). The second method, known as the 4 nearest neighbors approach, implements linear regression with the four nearest neighbors as predictors, while high-resolution observations serve as predictands (*Wilby and Wigley* 1997).

4. Analogs: this method looks for fields with similar conditions to the one being predicted in the reference large-scale historical dataset (at the coarser resolution). Once the date for the best past analog is identified, then the high-resolution reference field is retrieved as the predicted one. This approach is known as perfect prognosis because it assumes that the model has the same behaviour as the reference(*Wu et al.* 2012).

Due to these combinations, 37 methods have been compared for temperature and 42 for precipitation. These comparisons have been conducted with four metrics (two deterministic and two probabilistic. For deterministic predictions, we have used the Anomaly Correlation Coefficient (ACC; *Wilks* (2011)), which ranges from -1 (indicating total inverse correlation) to 1 (indicating perfect correlation) and the Root Mean Squared Error Skill Score (RMSSS; *Murphy* (1988)); whereas for probabilistic predictions, the Ranked Probability Skill Score (RPSS; *Wilks* (2011)) for tercile categories (i.e. below-normal, normal and above normal) and the Continuous Ranked Probability Skill Score (CRPSS; *Wilks* (2011)).

The skill scores have been chosen to assess the skill of the model in comparison to a reference forecast. Skill scores range from -infinite to 1, with positive values indicating better performance of the forecast than the reference forecast, meaning the opposite otherwise. In this study, the climatological forecast is used as the reference forecast because it is the most readily available for the users. For the deterministic evaluation, the climatological forecast is defined as no anomaly while, for the probabilistic evaluation, it is defined identical likelihood for all categories (i.e., probability of 33.3% for each tercile).

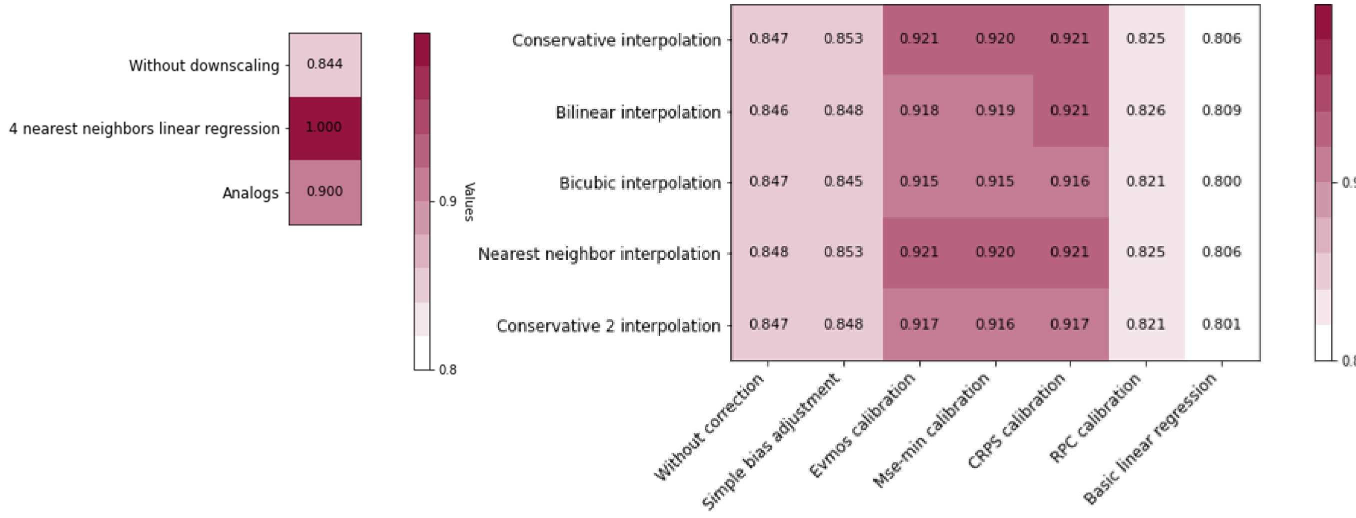
Regarding the statistical significance of the results, a one-sided t-test has been applied to assess if ACC values significantly differ from zero at the 95% confidence level (*Wilks* 2011). The significance of RMSSS and RPSS has been evaluated using the one-sided Random Walk Test applied to RMSE and RPS time series (*DelSole and Tippett* 2014). This test determines if the frequency of one forecast outperforming the other is significant at a 95% confidence level, focusing solely on the number of occurrences rather than the skill score value.

To summarize all the skill estimates obtained with the different metrics and facilitate the choice of the best overall approach, the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS; *Hwang et al.* (1981)) has been used. TOPSIS determines a relative score of the downscaling methods compared to the ideal value by normalizing all the considered metrics in a common ranking. The normalization of each metric is conducted on the basis of its ideal value, which is the highest value of the metric for that group of methods. For instance, if a downscaling technique obtains an average ACC value of 0.8, being the highest among all the other ACC values, it is considered the ideal value.

IV. RESULTS AND DISCUSSION

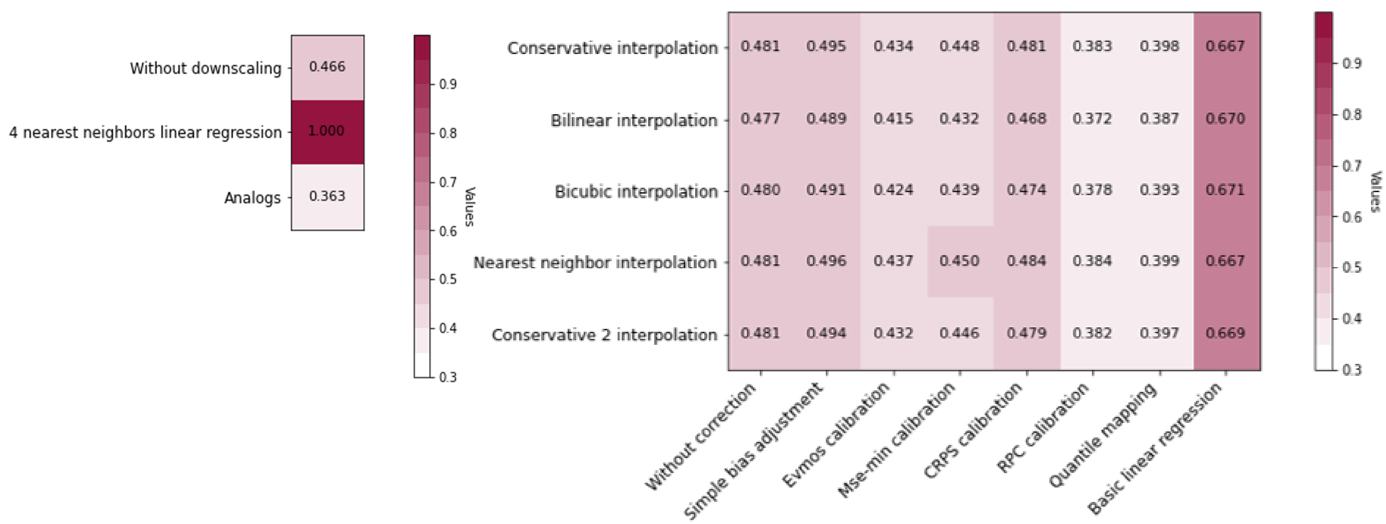
In this study we have evaluated the forecast quality of around 40 downscaling strategies (37 for temperature-based variables and 42 for precipitation) in the SADC region, for five different variables (i.e. mean temperature, mean precipitation, cold extremes, hot extremes and precipitation extremes) and two forecast periods (forecast year 1 and years 1-5), using four different verification metrics (ACC, RMSSS, RPSS, CRPSS) and the TOPSIS coefficient. However, only a set of results are presented in this paper due to space limitations: those for mean temperature and precipitation for forecast years 1-5 (showing TOPSIS and ACC in the main text, and the RPSS in the Annexes). The rest of the results are available in the R Shiny App (https://earth.bsc.es/shiny/smoreno/).

First, the results obtained using TOPSIS for tempera-



(a) (b)

Figure 1: TOPSIS coefficient results for mean temperature for all the downscaling methods. The five types of interpolations performed, along with the respective calibrations are displayed in (a). The results without downscaling and the two methods which do not employ any interpolation (4 nearest neighbors linear regression and analogs) are shown in (b).



(a) (b)

Figure 2: As Figure 1, but for precipitation.

ture and precipitation are presented in Figures 1 and 2, respectively. The higher skill obtained for temperature than precipitation showed in TOPSIS coefficients is in accordance with previous studies (*Delgado-Torres et al.* 2022, *Smith et al.* 2019).

Despite the discrepancies between variables, some similar patterns can be observed in the TOPSIS coefficients. For both variables, the values obtained for the same type of calibration, which appear in the same column of Figures 1b and 2b, are very similar for all interpolation methods, with slight differences in value across rows. These small differences indicate that the type of calibration applied carries more importance in the TOPSIS coefficient than the interpolation method employed. This is in consistent with the results obtained for each interpolation type using the same calibration, which show very small differences among them.

Turning to the spatial distribution of the ACC and percentage of grid points showing statistically significant correlation values for temperature predictions, they are

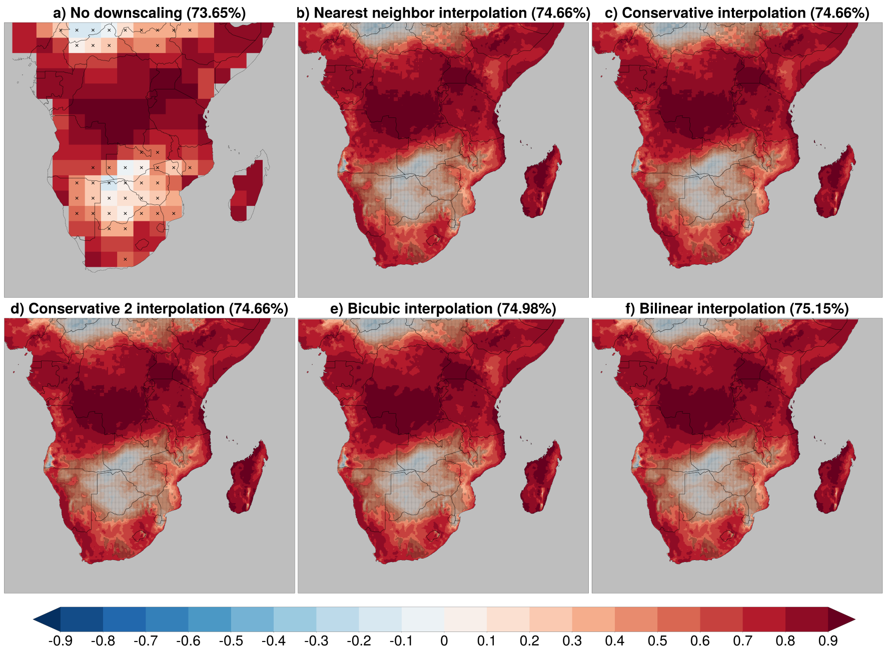


Figure 3: ACC obtained for temperature with no interpolation (a) and using different interpolation methods (b-f) for forecast years 1-5. The results are obtained comparing them with ERA5land reanalysis data. Additionally, the results obtained without downscaling is shown. The percentage of significant grid points is displayed in the subtitles. Points with no significant skill are shaded.

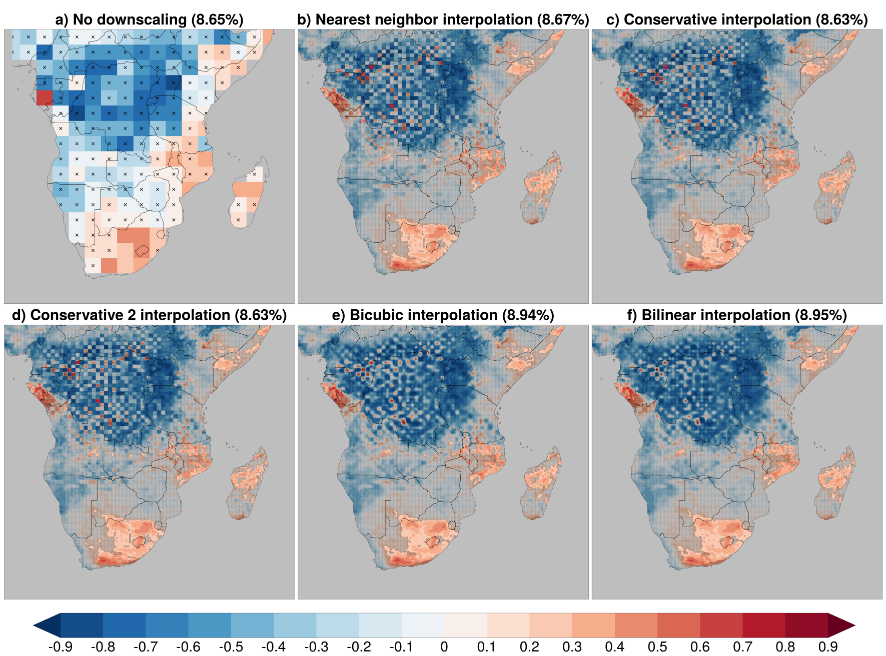


Figure 4: As Figure 3, but for precipitation

very similar for all interpolation methods (shown in Figure 3 ordered according to the TOPSIS value). The areas where the multi-model exhibits skill after applying the interpolation follow a spatial distribution similar to that observed without downscaling. However, the percentage of significant points for this particular metric and TOPSIS coefficients reveal that downscaling methods employing only interpolation on the observation grid show higher skill compared to the large-scale reference resolution.

This behavior observed for temperature interpolations is also evident in the case of precipitation. The differ-

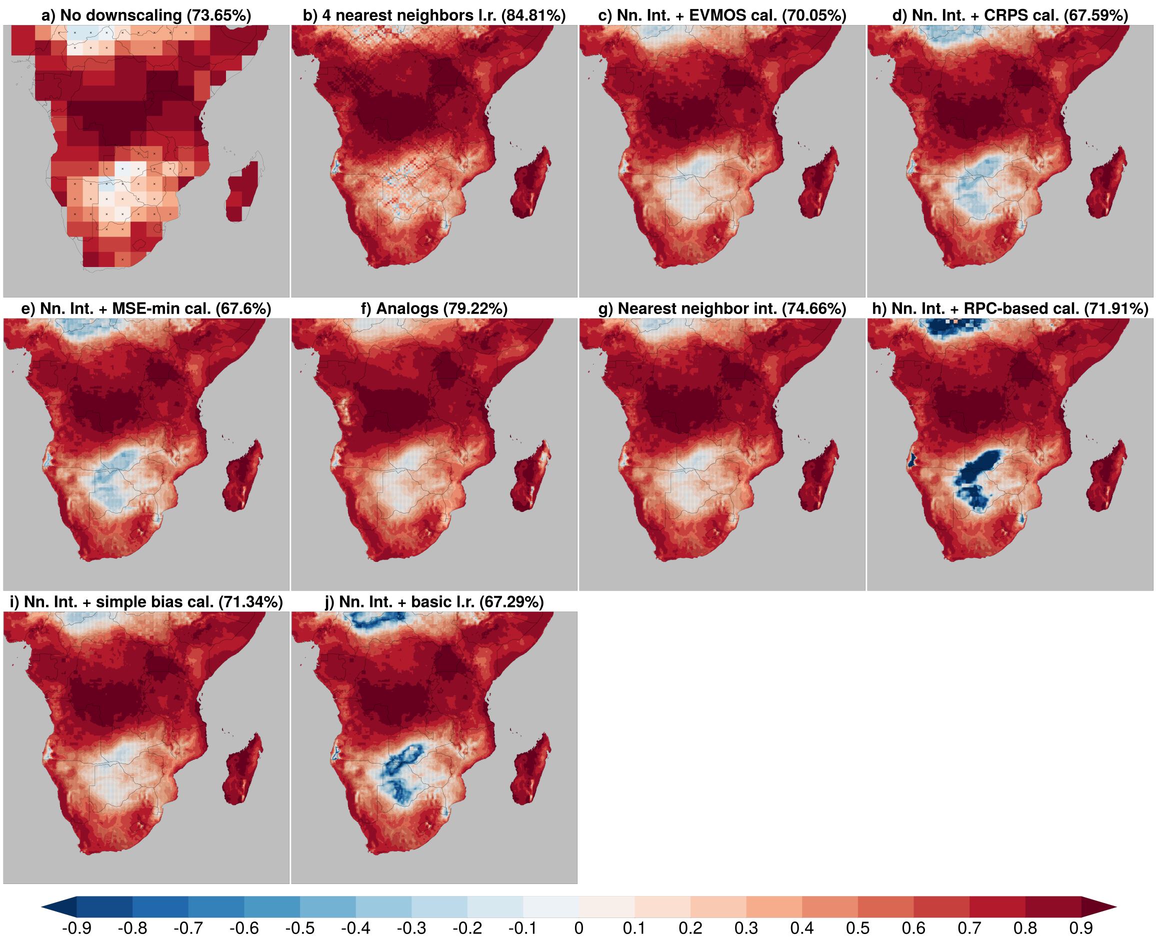


Figure 5: (a) ACC obtained for temperature for forecast years 1-5 with no downscaling; (b-j) using nearest neighbor interpolation for the downscaling methods. The reference dataset is ERA5land reanalysis. Additionally, the results obtained without downscaling are shown. The percentage of significant grid points is displayed in the subtitles. Points with no significant skill are shaded.

ent interpolation methods shown in Figure 4 exhibit a similar spatial distribution of areas with significant skill, as well as a comparable percentage of significant correlation values. Besides, Figure 2 shows that all of them obtain a higher TOPSIS coefficient than the reference values without downscaling.

The comparison of interpolation techniques (including the ones analyzed in this paper) has been already conducted in numerous works. For instance, *Hossain et al.* (2021) compared the five interpolation methods discussed in this study, among others, for monthly precipitation over Australia. Their results are in line with those presented here, indicating that all the interpolation methods exhibited similar skill and regional distribution. However, further work is needed to see whether these results are also applicable to other regions or variables.

Due to the small differences observed between interpolation methods, the results are presented solely based on the interpolation method that most often yields the best outcomes: the nearest neighbor method for both variables. Therefore, the visual comparison of the results from now onwards is only provided with this interpolation technique. It is worth noting, nevertheless, that although nearest neighbor interpolation may not provide the absolute best results for all the calibrations (e.g., for temperature RPC-based calibration, bilinear interpolation performs better than the nearest neighbor; Figure 1b), the observed differences in TOPSIS are minimal.

According to the TOPSIS coefficient values (Figures 1a and 2a), the 4 nearest neighbors linear regression method provides the highest skill estimates for both variables, yielding a perfect TOPSIS coefficient.

If we focus separately for each variable, there are also other methods that provide higher skill than the raw predictions for both variables. For temperature, for example, calibration-based methods adjusting the bias and the variance yield better results than predictions without any downscaling applied (with the CRPS, EVMOS, and MSE-min calibration methods providing the highest skill). Moreover, the analogs method has also delivered remarkably favorable outcomes in this context.

The ACC obtained for each method, filtered only for

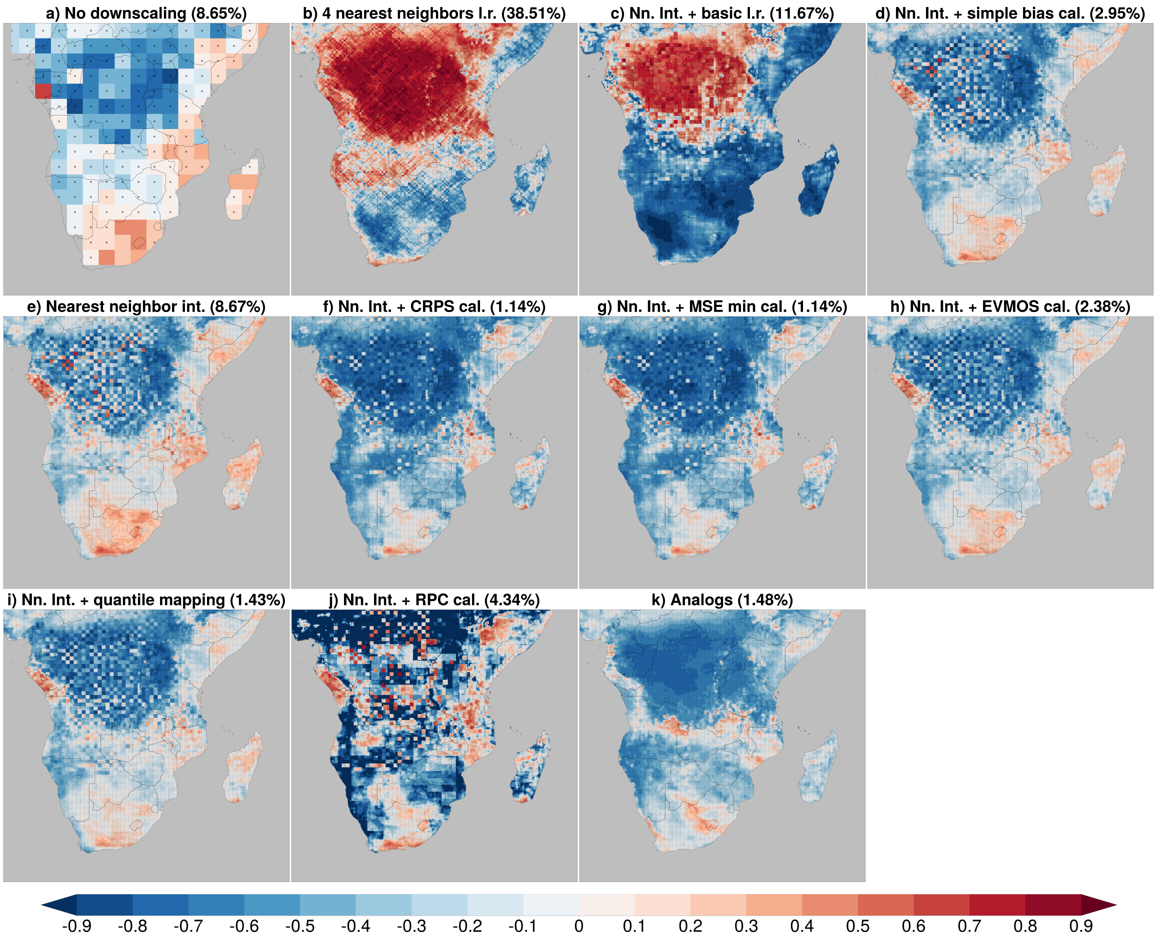


Figure 6: As Figure 5, but for precipitation.

the nearest neighbor interpolation, including analogs and 4 nearest neighbors linear regression along with the reference value without downscaling, are shown in Figure 5, ordered according to the obtained TOPSIS coefficient. Firstly, it is important to emphasize that, since TOPSIS combines information from the four metrics, the ranking according to TOPSIS in this figure may not correspond to the ranking of methods based solely on the ACC.

As for the interpolations, the skill distribution maps for ACC exhibit a spatial pattern similar to the reference map without downscaling. For this particular metric, all methods show high skill in the central part of the considered regions and nearly all coastal areas. Nonetheless, this is not the case for the inland regions in the southern and northern parts of the SADC, where all methods exhibit low prediction quality. This decrease of quality is even more pronounced for cases involving RPC-based calibration (Figure 5h), due to the nature of this method (*Eade et al.* 2014). RPC-based calibration focuses on correcting the signal-to-noise ratio for values that exhibit skill. When the original skill is not statistically significant, it directly replaces it with the climatology value, resulting in a metric score of -1. The same behaviour is manifested for skill distribution of the RPSS metric, shown in the Annexes in Figure S2.

Until now, only a limited number of studies have been conducted on downscaling methods of decadal climate predictions. Therefore, it is not straightforward to directly compare the results obtained in this study with any equivalent findings. Nonetheless, similar investigations have been carried out at the seasonal temporal scale, revealing consistent behavioral patterns. For instance, a work by *Manzanas et al.* (2018) for downscaling seasonal temperature predictions in Europe observed that the analogs and linear regression methods without interpolation added skill to the results without affecting their spatial distribution. Likewise, *Alfonso Hernanz* (2021) demonstrated that these two methods enhanced the skill of the raw model output for seasonal predictions in Spain. This behaviour is similar to the one obtained here, where the analogs method and the 4 nearest neighbors linear regression have obtained high TOPSIS values without showing any differential skill distribution at regional level (Figures 5b and 5f).

For precipitation predictions, many of the downscaling methods return lower skill than the non-downscaled predictions (Figure 2). However, in addition to the 4 nearest neighbors linear regression, some other techniques such as basic linear regression, simple bias adjustment and CRPS calibration provide higher skill than the raw predictions. In particular, the use of interpolation alone provides higher skill than when additionally applying calibrations. Nevertheless, calibration is a necessary step in a decision-making context as it improves the reliability and correct the statistical properties of the predictions for them to be more similar to the observations (*Doblas-Reyes et al.* 2005).

The results obtained for the ACC for precipitation with the different downscaling methods using only the nearest neighbor interpolation, the reference value without downscaling, and the two methods that do not use interpolation are displayed in Figure 6, ranked from best to worst. It can be seen that the spatial distribution of areas showing skill is very similar to that of the raw predictions for most of methods. Despite maintaining a similar spatial distribution, the original results already had limited skill throughout the region, and this skill is further reduced for almost all of the methods employed. This behaviour is also shown in Figure S2, where the RPSS distribution for precipitation is similar to the raw predictions for almost all the methods.

Nevertheless, in the two methods that show the highest skill (i.e. 4 nearest neighbors linear regression and interpolation with basic linear regression; Figures 6b and 6c, respectively), a particular behaviour is displayed: the central area of SADC region, which originally had negative skill before downscaling for both metrics (Figures 6a and S2a), shows a significantly positive skill. The exceptional performance of these methods in this certain area can be attributed to the ability of the downscaling method to correct specific characteristics of the predictions. Figure S3 in the Annexes shows the temporal evolution of precipitation anomalies in a specific location within this region (longitude 20°E, latitude 1°N), comparing the results obtained using the 4 nearest neighbors linear regression method with those without downscaling and with observed time series (for the individual models and the multi-model). The comparison reveals the remarkable performance of the downscaling method for the multi-model, overcoming the raw predictions.

In the case of the RPC-based calibration method, it yields remarkable results with highly negative values in many areas of the map (Figure 6j), reaching ACC values close to -1, which is due to the nature of this method as it replaces the original grid points that do not show skill with the climatology.

In line with temperature, there is limited research on downscaling for decadal climate predictions of precipitation. Nonetheless, a recent study conducted by *Paxian et al.* (2022) focused on downscaling decadal predictions of precipitation in Germany using a combination of analogs and linear regression, showing promising outcomes. Consistent with these findings, our study also highlights the effectiveness of linear regression-based methods, which have exhibited notable skill improvements when applying downscaling. Other studies have been conducted on downscaling seasonal predictions of precipitation. For example, *Tabari et al.* (2021) have explored the use of simple bias adjustment as a downscaling method, yielding comparable outcomes to those obtained in our study.

It is worth noting that there are potential sources of error associated with statistical downscaling methods. First, these methods assume the stationarity of the relationship between predictors and predictands. However, in the context of climate change, non-stationarity behaviours become a significant concern (*Fowler et al.* 2007). Second, the accuracy of the downscaled results depends on the quality and limitations of climate models used for deriving the large-scale predictors. If the models produce inaccurate predictions, it can impact the quality of the downscaled outputs. Therefore, careful selection of variables and the choice of an appropriate statistical methodology play a vital role. Simple linear models can yield misleading conclusions by generating artificial linear relationships between predictors and predictands.

V. CONCLUSIONS

In this work we have carried out a forecast quality assessment of downscaled decadal predictions for the SADC region using all available decadal predictions from the forecast systems contributing to CMIP6/DCPP-A. This has been achieved upon a multi-model approach combining the predictions from 13 different forecast systems. This evaluation has been applied for the forecast years 1 and 1-5 for two essential climate variables: near-surface air temperature and precipitation, and three extreme indices based on daily minimum and maximum temperature and precipitation (TN10p, TX90p and R95p). The forecast quality has been assessed from two viewpoints: one, deterministic, through the ACC and RMSSS and the other, probabilistic, with the RPSS and CRPSS. On the other hand, the raw multi-model has been set as the benchmark to help us identify the potential improvements the downscaling methods could bring.

The results obtained vary depending on the variable, region and forecast period considered. However, both variables (temperature and precipitation) show similar behaviours for some characteristics. In the two cases, for example, the choice of the interpolation method does not remarkably impact the forecast quality and the ranking depends on the calibration or linear regression techniques. Besides, regarding the selection of the best method through the TOPSIS coefficient, 4 nearest neighbors linear regression provides the highest skill results for both variables.

In addition to this method, for temperature, although the overall skill distribution across the region generally remains similar to the raw forecasts, it is noteworthy that some areas show a moderate enhancement in skill, presumably due to the application of specific downscaling techniques. These improvements in skill seem to be possible when adjustments are made to both the mean value and the variance in the model output, when the CRPS calibration is applied, or when the Mean Squared Error-minimization (MSE-min) method is employed. Additionally, the use of the analogs method has also shown potential for achieving reasonably high forecast quality, although further exploration is necessary to confirm these preliminary observations.

When considering precipitation predictions, both the 4 nearest neighbors method and the basic linear regression have shown a good transition from the coarse resolution to the finer one, maintaining the performance or even increasing it in some areas and forecast periods. This is also true, but not equally general, for simple bias correction and interpolation-only approaches.

In summary, this research establishes an initial benchmark for future research in the SADC area, fostering further exploration and analysis in the field of multi-model downscaling of decadal predictions. In terms of future work, we aim to extend the boundaries of this research by examining additional downscaling techniques and incorporating an expanded range of variables and forecast periods (some of these are already included in the R Shiny App, https://earth.bsc.es/shiny/smoreno/). The ultimate objective of these efforts is the publication of a manuscript based on this study.

ACKNOWLEDGMENTS

This study has been supported by the FOCUS-Africa project, founded by the European Union under Horizon 2020 Programme (grant agreement 869575). I wish to express my sincere gratitude to my tutors Carlos Delgado and Raül Marcos, for their guidance and support throughout this research, and to the Earth Sciences Department of the Barcelona Supercomputing Center for the collaboration and the opportunity to work alongside the team members. Special thanks to my collegues Jaume Ramon and Eren Duzenli for their technical support. The authors also thank the modelling centres for making predictions available through the ESGF node (https://esgf-node.llnl.gov/search/cmip6/).

REFERENCES

Alfonso Hernanz, M. D. P. R.-C. M. A. P.-S. E. R.-C., Juan Andrés García-Valero, Evaluation of statistical downscaling methods for climate change projections over spain: Present conditions with perfect predictors, *International Journal of Climatology*, *42*(2), 762–776, doi: https://doi.org/10.1002/joc.7271, 2021.

Bethke, I., et al., Norcpm1 and its contribution to cmip6 dcpp, *Geoscientific Model Development*, *14*(11), 7073–7116, 2021.

Boer, G. J., Decadal potential predictability of twenty-first century climate, *Climate dynamics*, *36*, 1119–1133, doi: 10.1007/s00382-010-0747-9, 2011.

Boer, G. J., et al., The decadal climate prediction project (dcpp) contribution to cmip6, *Geoscientific Model Development*, *9*(10), 3751–3777, doi:10.5194/gmd-9-3751-2016, 2016.

Boucher, O., et al., Presentation and evaluation of the ipsl-cm6a-lr climate model, *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002,010, doi: 10.1029/2019MS002010, 2020.

Bruno Soares, M., M. Alexander, and S. Dessai, Sectoral use of climate information in europe: A synoptic overview, *Climate Services*, *9*, 5–20, doi: https://doi.org/10.1016/j.cliser.2017.06.001, climate services in practice: what we learnt from EUPORIAS, 2018.

Delgado-Torres, C., et al., Multi-model forecast quality assessment of cmip6 decadal predictions, *Journal of Climate*, *35*(13), 4363–4382, doi:10.1175/JCLI-D-21-0811.1, 2022.

DelSole, T., and M. K. Tippett, Comparing forecast skill, *Monthly Weather Review*, *142*(12), 4658–4678, doi: 10.1175/MWR-D-14-00045.1, 2014.

Doblas-Reyes, F., I. Andreu-Burillo, Y. Chikamoto, J. García-Serrano, V. Guemas, M. Kimoto, T. Mochizuki, L. Rodrigues, and G. Van Oldenborgh, Initialized near-term regional climate change prediction, *Nature communications*, *4*(1), 1715, doi:10.1038/ncomms2704, 2013.

Doblas-Reyes, F. J., R. Hagedorn, and T. Palmer, The rationale behind the success of multi-model ensembles in seasonal forecasting—ii. calibration and combination, *Tellus A: Dynamic Meteorology and Oceanography*, *57*(3), 234–252, doi:10.1111/j.1600-0870.2005.00103.x, 2005.

Döscher, R., et al., The ec-earth3 earth system model for the coupled model intercomparison project 6, *Geoscientific Model Development*, *15*(7), 2973–3020, doi:10.5194/gmd-15-2973-2022, 2022.

Dunstone, N., et al., Skilful interannual climate prediction from two large initialised model ensembles, *Environmental Research Letters*, *15*(9), 094,083, doi:10.1088/1748-9326/ab9f7d, 2020.

Eade, R., D. Smith, A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson, Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?, *Geophysical Research Letters*, *41*(15), 5620–5628, doi:https://doi.org/10.1002/2014GL061146, 2014.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization, *Geoscientific Model Development*, *9*(5), 1937–1958, doi:10.5194/gmd-9-1937-2016, 2016.

Fang, G., J. Yang, Y. Chen, and C. Zammit, Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in china, *Hydrology and Earth System Sciences*, *19*(6), 2547–2559, doi: 10.5194/hess-19-2547-2015, 2015.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi, Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *27*(12), 1547–1578, doi: https://doi.org/10.1002/joc.1556, 2007.

Gneiting, T., and A. Raftery, Atmospheric science — weather forecasting with ensemble methods, *Science (New York, N.Y.)*, *310*, 248–9, doi:10.1126/science.1115255, 2005.

Gutierrez, J., J. Bedia, R. Benestad, and C. Pagé, Local predictions based on statistical and dynamical downscaling, *Technical Report 308378*, 2013.

Hazeleger, W., V. Guemas, B. Wouters, S. Corti, I. Andreu-Burillo, F. J. Doblas-Reyes, K. Wyser, and M. Caian, Multiyear climate predictions using two initialization strategies, *Geophysical Research Letters*, *40*(9), 1794–1798, doi: 10.1002/grl.50355, 2013.

Hossain, M. M., N. Garg, A. F. Anwar, and M. Prakash, Comparing spatial interpolation methods for cmip5 monthly precipitation at catchment scale, *Sustainable water resources development and management. Excel India Publishers, Aurangabad*, *285*, 2021.

Hwang, C.-L., K. Yoon, C.-L. Hwang, and K. Yoon, Methods for multiple attribute decision making, *Multiple attribute decision making: methods and applications a state-of-the-art survey*, pp. 58–191, doi:10.1007/978-3-319-91648-4˙1, 1981.

Keller, A. A., K. L. Garner, N. Rao, E. Knipping, and J. Thomas, Downscaling approaches of climate change projections for watershed modeling: Review of theoretical and practical considerations, *PLOS Water*, *1*(9), e0000,046, doi: 10.1371/journal.pwat.0000046, 2022.

Leung, L. R., A. F. Hamlet, D. P. Lettenmaier, and A. Kumar, Simulations of the enso hydroclimate signals in the pacific northwest columbia river basin, *Bulletin of the American Meteorological Society*, *80*(11), 2313 – 2330, doi:https://doi.org/10.1175/1520-0477(1999)080¡2313:SOTEHS¿2.0.CO;2, 1999.

Manzanas, R., J. Gutiérrez, E.-C. S. Fernández, J.and Van Meigjaard, M. Magariño, A. Cofiño, and S. Herrera, Dynamical and statistical downscaling of seasonal temperature forecasts in europe: Added value for user applications, *Climate Services*, *9*(9), 44–56, doi: 10.1016/j.cliser.2017.06.004, 2018.

Muñoz Sabater, J., et al., Era5-land: a state-of-the-art global reanalysis dataset for land applications, *Earth System Science Data*, *13*(9), 4349–4383, doi:10.5194/essd-13-4349-2021, 2021.

Müller, W. A., et al., A higher-resolution version of the max planck institute earth system model (mpi-esm1. 2-hr), *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383–1413, doi:10.1029/2017MS001217, 2018.

Murphy, A. H., Skill scores based on the mean square error and their relationships to the correlation coefficient, *Monthly weather review*, *116*(12), 2417–2424, doi:10.1175/1520-0493(1988)116¡2417:SSBOTM¿2.0.CO;2, 1988.

Nicolì, D., A. Bellucci, P. Ruggieri, P. J. Athanasiadis, S. Materia, D. Peano, G. Fedele, R. Hénin, and S. Gualdi, The euro-mediterranean center on climate change (cmcc) decadal prediction system, *Geoscientific Model Development*, *16*(1), 179–197, doi:10.5194/gmd-16-179-2023, 2023.

Paxian, A., M. Ziese, F. Kreienkamp, K. Pankatz, S. Brand, A. Pasternack, H. Pohlmann, K. Modali, and B. Früh, User-oriented global predictions of the gpcc drought index for the next decade, *Meteorologische Zeitschrift*, *28*, 3–21, 2019.

Paxian, A., et al., High-resolution decadal drought predictions for german water boards: A case study for the wupper catchment, *Frontiers in Climate*, *4*, doi: 10.3389/fclim.2022.867814, 2022.

Schoof, J. T., Statistical downscaling in climatology, *Geography Compass*, *7*(4), 249–265, doi:10.1111/gec3.12036, 2013.

Sellar, A. A., et al., Implementation of uk earth system models for cmip6, *Journal of Advances in Modeling Earth Systems*, *12*(4), e2019MS001,946, 2020.

Slingo, J., and T. Palmer, Uncertainty in weather and climate prediction, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1956), 4751–4767, doi:10.1098/rsta.2011.0161, 2011.

Smith, D., et al., Robust skill of decadal climate predictions, *Npj Climate and Atmospheric Science*, *2*(1), 13, doi: 10.1038/s41612-019-0071-y, 2019.

Smith, D. M., R. Eade, and H. Pohlmann, A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction, *Climate Dynamics*, *41*, 3325–3338, doi: 10.1007/s00382-013-1683-2, 2013.

Solaraju-Murali, B., N. Gonzalez-Reviriego, L.-P. Caron, A. Ceglar, A. Toreti, M. Zampieri, P.-A. Bretonnière, M. Samsó Cabré, and F. J. Doblas-Reyes, Multi-annual prediction of drought and heat stress to support decision making in the wheat sector, *npj Climate and Atmospheric Science*, *4*(1), 34, 2021.

Swart, N. C., et al., The canadian earth system model version 5 (canesm5. 0.3), *Geoscientific Model Development*, *12*(11), 4823–4873, doi:10.5194/gmd-12-4823-2019, 2019.

Tabari, H., S. M. Paz, D. Buekenhout, and P. Willems, Comparison of statistical downscaling methods for climate change impact analysis on precipitation-driven drought, *Hydrology and Earth System Sciences*, *25*(6), 3493–3517, doi: 10.5194/hess-25-3493-2021, 2021.

Tatebe, H., et al., Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in miroc6, *Geoscientific Model Development*, *12*(7), 2727–2765, 2019.

Van Schaeybroeck, B., and S. Vannitsem, Post-processing through linear regression, *Nonlinear Processes in Geophysics*, *18*(2), 147–160, doi:10.5194/npg-18-147-2011, 2011.

Wilby, R. L., and T. M. Wigley, Downscaling general circulation model output: a review of methods and limitations, *Progress in physical geography*, *21*(4), 530–548, doi: 10.1177/030913339702100403, 1997.

Wilks, D. S., *Statistical methods in the atmospheric sciences*, vol. 100, Academic press, doi:10.2307/2669579, 2011.

Wu, T., et al., The beijing climate center climate system model (bcc-csm):¡? xmltex\break?¿ the main progress from cmip5 to cmip6, *Geoscientific Model Development*, *12*(4), 1573–1600, doi:10.5194/gmd-12-1573-2019, 2019.

Wu, W., et al., Statistical downscaling of climate forecast system seasonal predictions for the southeastern mediterranean, *Atmospheric Research*, *118*, 346–356, doi: 10.1016/j.atmosres.2012.07.019, 2012.

Yukimoto, S., et al., The meteorological research institute earth system model version 2.0, mri-esm2. 0: Description and basic evaluation of the physical component, *Journal of the Meteorological Society of Japan. Ser. II*, *97*(5), 931–965, doi:10.2151/jmsj.2019-051, 2019.

Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, Indices for monitoring changes in extremes based on daily temperature and precipitation data, *WIREs Climate Change*, *2*(6), 851–870, doi:https://doi.org/10.1002/wcc.147, 2011.

APPENDIX

Forecast system	Institution	DCPP members	Spatial resolution	Initialization month	Reference
BCC-CSM2-MR	BCC	8	1.125°x 1.125°	January	(*Wu et al.* 2019)
CanESM5	CCCma	20	2.8°x 2.8°	January	(*Swart et al.* 2019)
CMCC-CM2-SR5	CMCC	10	0.9°x 1.25°	November	(*Nicolì et al.* 2023)
EC.Earth-i1	BSC	10	0.7°x 0.7°	November	(*Döscher et al.* 2022)
EC-Earth-i2	SMHI/DMI	5	0.7°x 0.7°	November	(*Döscher et al.* 2022)
EC-Earth-i4	BSC	10	0.7°x 0.7°	November	(*Döscher et al.* 2022)
HadGEM3-GC3.1-MM	MOHC	10	0.55°x 0.83°	November	(*Sellar et al.* 2020)
IPSL-CM6A-LR	IPSL	10	1.25 °x 2.5 °	January	(*Boucher et al.* 2020)
MIROC6	MIROC	10	1.4°x 1.4°	November	(*Tatebe et al.* 2019)
MPI-ESM1.2-HR	DWD	10	0.9°x 0.9°	November	(*Müller et al.* 2018)
MRI-ESM2-0	MRI	10	1.125°x 1.125°	November	(*Yukimoto et al.* 2019)
NorCPM1-i1	NCC	10	1.9°x 2.5°	October	(*Bethke et al.* 2021)
NorCPM1-i2	NCC	10	1.9°x 2.5°	October	(*Bethke et al.* 2021)

Table S1: Forecast systems that contribute to the DCPP-A component of CMIP6 and their specifications, including the available simulations at the time of the study and the atmospheric grid's spatial resolution (latitude x longitude).

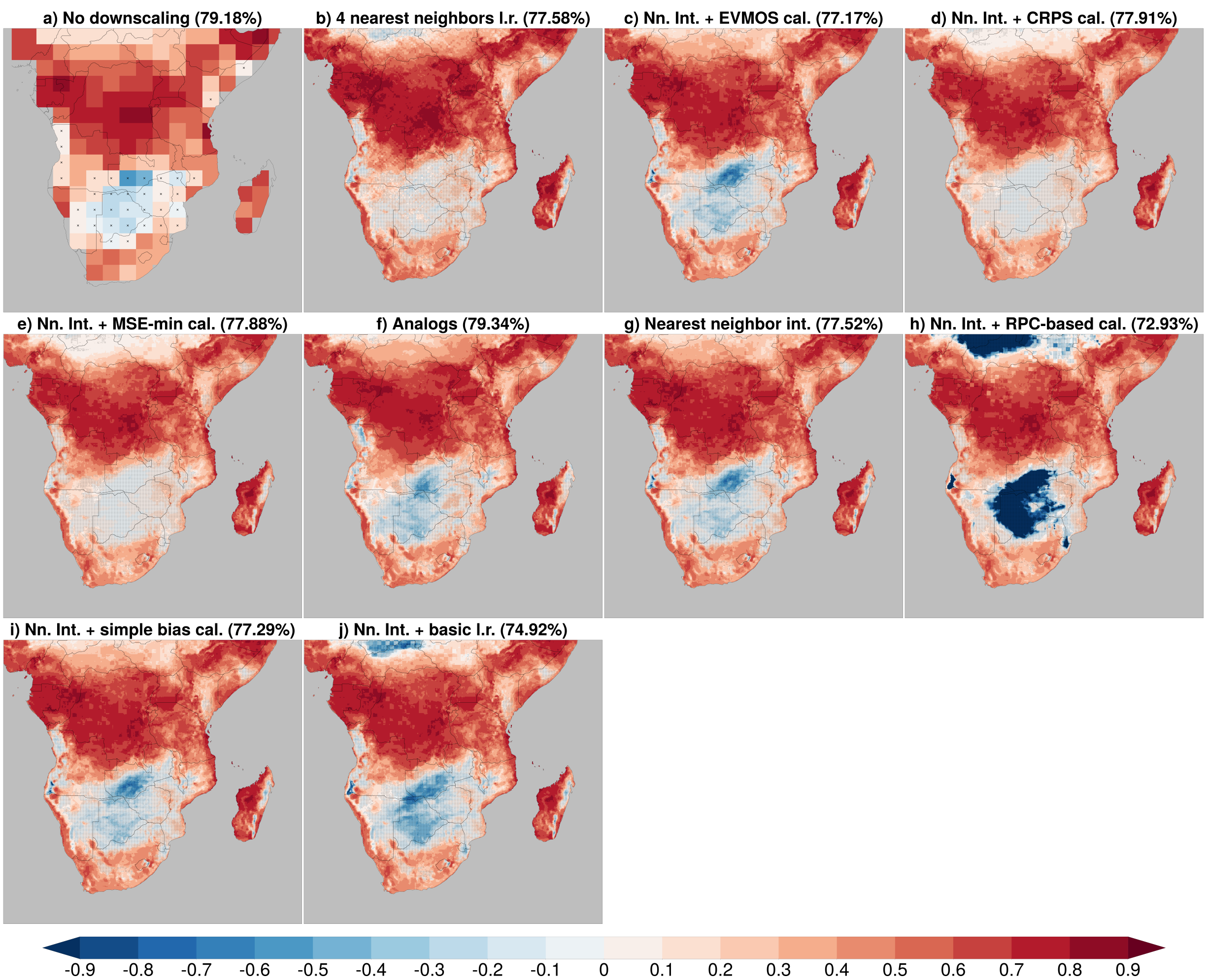


Fig. S1: As Figure 5, but for the RPSS.

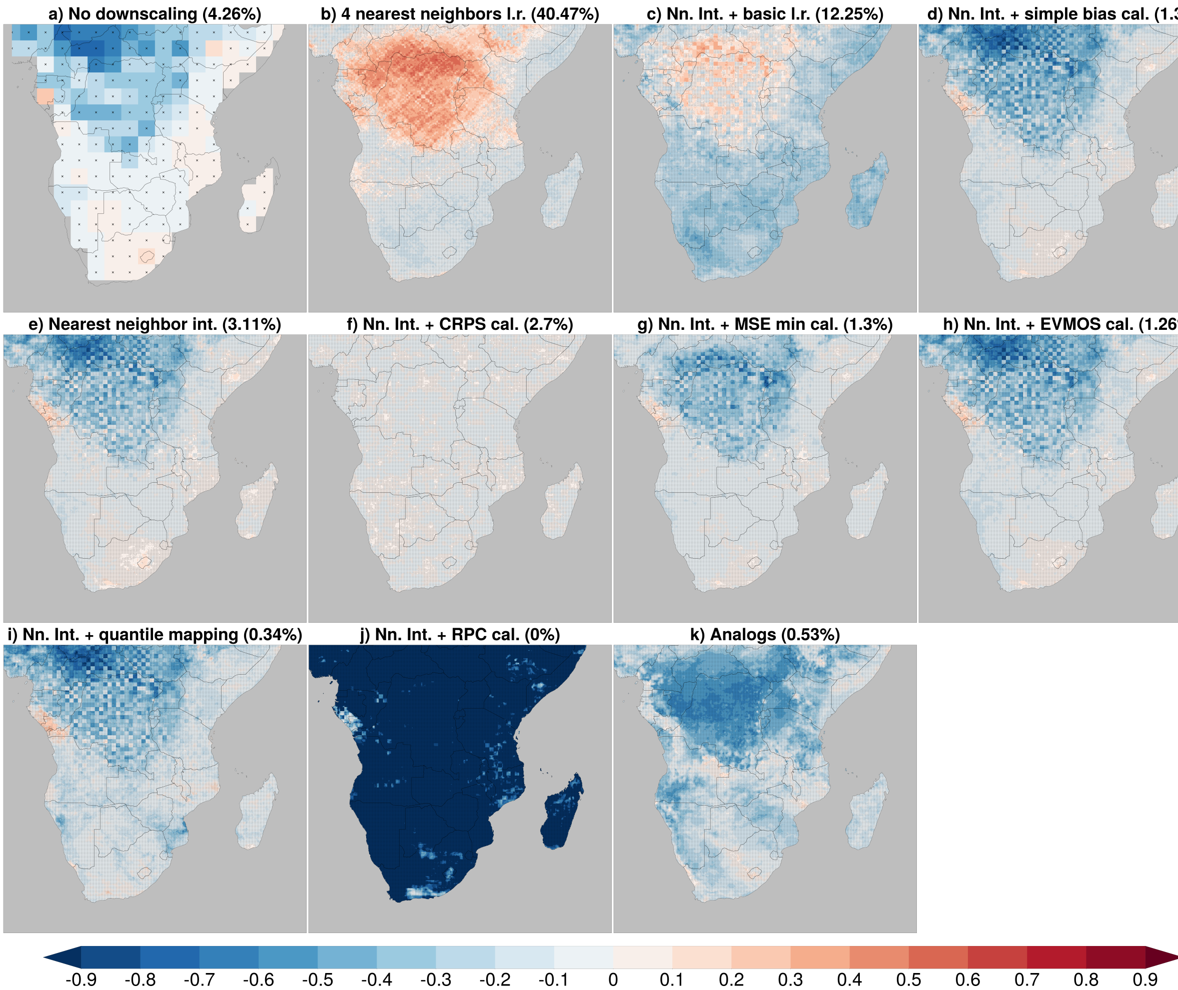


Fig. S2: As Figure 6, but for the RPSS.

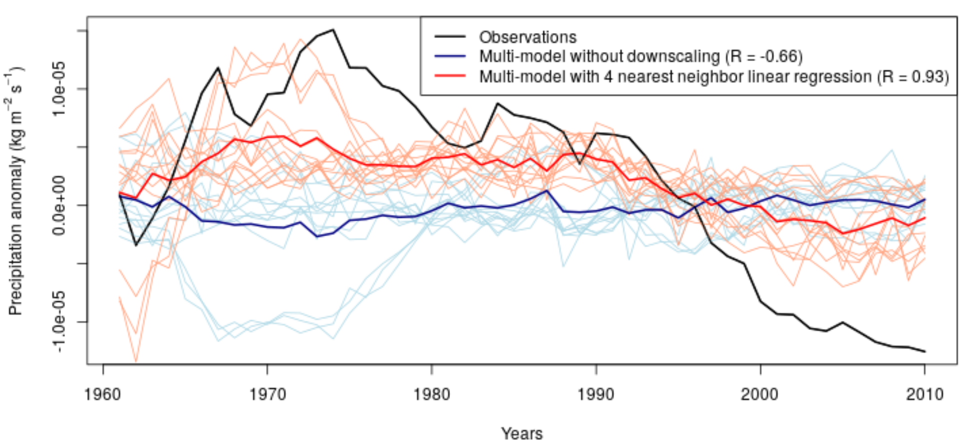


Fig. S3: Precipitation anomaly temporal evolution from observations, downscaled data using 4 nearest neighbors linear regression and data without downscaling for longitude 20°E and latitude 1°N. Thick lines represent the multi-model values while thin lines represent each model values individually.