FOCUS-Africa

Research and Innovation Action (RIA)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869575

Start date : 2020-09-01 Duration : 48 Months

Report on theverification of seasonalforecasts and thecharacterization ofclimate projections anddeca predictions

Authors : Dr. Marcos-Matamoros RAüL (BSC-CNS), Raül Marcos?Matamoros, Chih Chung?Chou, Carlos Delgado, Scott?Burgan, Erasmo Buonomo, Marcello? Pettita, Laura Trentini, Chris Lennard, Piotr Wolski, Lisa van Ardenne, Peter Marsh, Chris Jack, François Engelbrecht, Jonathan Padavatan, Jessica Steinkof, Matteo Dell?Acqua, Robel Takele, Vaileth Jonas, Alfred Kondowe, Esther Verena Jansen, Andre Kamga, Sebastian Grey and Roberta Boscolo

FOCUS-Africa - Contract Number: 869575

Project officer: Anna-Natasa ASIK

Document title	Report on theverification of seasonalforecasts and thecharacterization ofclimate projections anddecadal predictions
Author(s)	Dr. Marcos-Matamoros RAüL, Raül Marcos?Matamoros, Chih Chung?Chou, Carlos Delgado, Scott?Burgan, Erasmo Buonomo, Marcello? Pettita, Laura Trentini, Chris Lennard, Piotr Wolski, Lisa van Ardenne, Peter Marsh, Chris Jack, François Engelbrecht, Jonathan Padavatan, Jessica Steinkof, Matteo Dell?Acqua, Robel Takele, Vaileth Jonas, Alfred Kondowe, Esther Verena Jansen, Andre Kamga, Sebastian Grey and Roberta Boscolo
Number of pages	147
Document type	Deliverable
Work Package	WP4
Document number	D4.1
Issued by	BSC-CNS
Date of completion	2022-11-03 15:08:06
Dissemination level	Public

Summary

This report accounts for deliverable 4.1, 'Report on the verification of seasonal forecasts and the characterization of climate projections and decadal predictions??. It contains the work performed in WP4 tasks 4.1, 'Seasonal forecast quality assessment' and 4.2, 'Climate projections and decadal assessment', during the period between M1 and M24. It is mainly centered on the methodologies implemented to postprocess and verify the seasonal, decadal, and climate projection data that will be used in the co?development of climate services for the FOCUS?Africa case studies. The first part of this deliverable focuses on task 4.1, 'Seasonal forecast quality assessment', and it contains: (i) the verification of raw and bias corrected ECVs from the seasonal forecasting models included in the C3S Data Store as well as a comparison of different bias correction approaches; (ii) the verification of a specifically tailored bias correction method for extreme ECV values in the South African Development Community (SADC); (iii) the review of current seasonal methodology at SARCOF and SIOCOF as part of RCOF session reports (developed in depth in deliverable 7.1); and (iv) the review on the available products for the onset of the rainy season in Tanzania. The second part includes the results from task 4.2: 'Climate projections and decadal assessment'. It reports the (i) assessment of decadal skill on ECVs and of the impact of different bias correction approaches on decadal predictions; (ii) the assessment of projection skills with a special focus on variables related to cereals and legume cropping; (iii) climate projections selection for case studies by comparing high?resolution RCM data with the GCMs in synergy with WP3; and (iv) statistical analysis of the CMIP6 climate projections with a special focus on extremes considering CORDEX?core regions and specific ensembles. The overall analysis of the three climate prediction time scales (seasonal, decadal, and climate projection) through different post?processing and verification strategies (i.e. bias correction) provides a thorough characterization of the current landscape of available approaches for the pilot case studies. Additionally, this review provides the needed foothold to move forward and undertake future tasks 4.3, 'Implementation of multi?model and downscaling for seasonal forecasts', 4.4, 'Implementation of multi?model and downscaling for climate projections and decadal predictions', and 4.5, 'Deri...

Approval

Date	By
2022-11-03 15:28:20	Dr. Marcos-Matamoros RAüL (BSC-CNS)
2022-11-12 15:56:05	Mrs. Roberta BOSCOLO (WMO)

FOCUS-AFRICA

Full-value chain Optimised Climate User-centric Services for Southern Africa

Report on the verification of seasonal forecasts and the characterization of climate projections and decadal predictions

Deliverable D4.1

Lead Beneficiary: Barcelona Supercomputing Centre (BSC)

24/2022

Raül Marcos-Matamoros[1], Chih Chung-Chou[1], Carlos Delgado[1], Scott-Burgan[2], Erasmo Buonomo[2], Marcello-Pettita[3], Laura Trentini[3], Chris Lennard[4], Piotr Wolski[4], Lisa van Ardenne[4], Peter Marsh[4], Chris Jack[4], François Engelbrecht[5], Jonathan Padavatan[5], Jessica Steinkof[5], Matteo Dell'Acqua[6], Robel Takele[6], Vaileth Jonas[7], Alfred Kondowe[7], Esther Verena Jansen[8], Andre Kamga[8], Sebastian Grey[8] and Roberta Boscolo[9]

[1] *Barcelona Supercomputing Centre (BSC)*
[2] *Met Office*
[3] *Amigo*
[4] *University of Cape Town (UCT)*
[5] *University of Witwatersrand (WITS)*
[6] *Scuola Superiore Sant'Anna (SSSA)*
[7] *Tanzania Meteorological Agency (TMA)*
[8] *African Centre of Meteorological Applications for Development (ACMAD)*
[9] *World Meteorological Organization (WMO)*

This project has received funding from the European Commission's Horizon 2020 Research and Innovation Programme. The content in this presentation reflects only the author(s)'s views. The European Commission is not responsible for any use that may be made of the information it contains.

www.focus-africaproject.eu

Document Information

Grant Agreement: 869575

Project Title: Full-value chain Optimised Climate User-centric Services for Southern Africa

Project Acronym: Focus-Africa

Project Start Date: 1 September 2020

Related Work Package: WP 4: Development of Methods and Tools

Lead Organisation: BSC

Submission date: 31 October 2022

History

Date	Submitted by	Reviewed by	Version (Notes)
25/10/2022	BSC		

About FOCUS-Africa

FOCUS-Africa – Full-value chain Optimised Climate User-centric Services for Southern Africa – is developing sustainable tailored climate services in the Southern African Development Community (SADC) region for four sectors: agriculture and food security, water, energy and infrastructure.

It will pilot eight case studies in six countries involving a wide range of end-uses to illustrate how the application of new climate forecasts, projections, resources from Copernicus, GFCS and other relevant products can maximise socio-economic benefits in the Southern Africa region and potentially in the whole of Africa.

Led by WMO, it gathers 14 partners across Africa and Europe jointly committed to addressing the recurring sustainability and exploitation challenge of climate services in Africa over a period of 48 months.

For more information visit: www.focus-africaproject.eu

Coordinator Contact
Roberta Boscolo | Climate & Energy Scientific Officer
Applied Climate Services Division
Services Department
World Meteorological Organization (WMO)
CP 2300, 1211 Geneva, Switzerland
Email: rboscolo@wmo.int

Table of Contents

Tables

Figures

Executive Summary

This report accounts for deliverable 4.1, 'Report on the verification of seasonal forecasts and the characterization of climate projections and decadal predictions''. It contains the work performed in WP4 tasks 4.1, 'Seasonal forecast quality assessment' and 4.2, 'Climate projections and decadal assessment', during the period between M1 and M24. It is mainly centered on the methodologies implemented to postprocess and verify the seasonal, decadal, and climate projection data that will be used in the co-development of climate services for the FOCUS-Africa case studies.

The first part of this deliverable focuses on task 4.1, 'Seasonal forecast quality assessment', and it contains: (i) the verification of raw and bias corrected ECVs from the seasonal forecasting models included in the C3S Data Store as well as a comparison of different bias correction approaches; (ii) the verification of a specifically tailored bias correction method for extreme ECV values in the South African Development Community (SADC); (iii) the review of current seasonal methodology at SARCOF and SIOCOF as part of RCOF session reports (developed in depth in deliverable 7.1); and (iv) the review on the available products for the onset of the rainy season in Tanzania.

The second part includes the results from task 4.2: 'Climate projections and decadal assessment'. It reports the (i) assessment of decadal skill on ECVs and of the impact of different bias correction approaches on decadal predictions; (ii) the assessment of projection skills with a special focus on variables related to cereals and legume cropping; (iii) climate projections selection for case studies by comparing high-resolution RCM data with the GCMs in synergy with WP3; and (iv) statistical analysis of the CMIP6 climate projections with a special focus on extremes considering CORDEX-core regions and specific ensembles.

The overall analysis of the three climate prediction time scales (seasonal, decadal, and climate projection) through different post-processing and verification strategies (i.e. bias correction) provides a thorough characterization of the current landscape of available approaches for the pilot case studies. Additionally, this review provides the needed foothold to move forward and undertake future tasks 4.3, 'Implementation of multi-model and downscaling for seasonal forecasts', 4.4, 'Implementation of multi-model and downscaling for climate projections and decadal predictions', and 4.5, 'Derived products using seasonal forecasts, climate projections and / or decadal predictions'. In this way it completes the first stage towards the provision of the most actionable information for the FOCUS-Africa case studies.

Keywords

Seasonal forecast, decadal predictions, climate projections, bias correction, quality assessment

1 Introduction

FOCUS-Africa aims to create innovative and sustainable climate services in four sectors of the Southern African Development Community (SADC): agriculture, food security, energy, and infrastructure. In this framework, Work Package 4 (WP4) seeks to promote the systematic use of advanced methods for the integration of seasonal, decadal and climate projections in the co-development of the eight pilot climate services that are being built within the project. These methods comprise topics such as skill assessment, downscaling, bias adjustment as well as multi-model construction and indicator tailoring. More specifically, the actions included in this deliverable involve the WP4 objectives: (i) improvement of the understanding of the seasonal predictability of the Essential Climate Variables, ECV (ii) improvement of the forecast performance through the application of bias correction approaches and, (iii) assessment of the long-term influences of climate in the areas of the case studies.

These goals have been fulfilled through the work performed within tasks 4.1 ('Seasonal forecast quality assessment') and 4.2 ('Climate projections and decadal assessment'). Regarding task 4.1, the estimation of the seasonal forecast quality based on past performance and verification metrics allows quantifying its potential benefit compared with other approaches. This is achieved by systematically collating the forecasts to a reference to assess their overall matching. Additionally, since the forecast systems are affected by biases, a range of methods have been applied to obtain similar statistical properties as found in the reference variables. This framework seeks to provide end-users with the information needed to understand which strategies would better fit their interests (important in WP5, WP6 and WP7). In this regard, with the objective of providing information that could be operationalised in the future, the seasonal prediction systems employed come from the Copernicus Climate Data Store (CCDS).

Moving forward to task 4.2, decadal and climate time-scales offer a unique framework for end-user exploration of the possible future evolutions of the climate systems. In this task, simulations from the state-of-the-art Coupled Model Intercomparison Project Phase 6 (CMIP6 and available CORDEX data) have been assessed and analysed to establish the potential changes that the FOCUS-Africa areas will experience in the coming decades. This assessment has been carried out in coordination with the complementary actions from WP3. Moreover, the bias correction methods that could be applied at these time-scales have also been evaluated in views of their potential application in the case studies.

In this framework, the organisation of the deliverable is as follows: after the introduction, three more sections account for the outcomes obtained for each of the three types of climate information available: seasonal forecasts, decadal predictions and climate projections. Each time-scale includes the reporting of the different subtasks carried out in 4.1 and 4.2. Finally, the conclusions section summarizes the results obtained.

2 Seasonal forecasts

2.1 Verification of raw and bias adjusted seasonal forecasts

Seasonal predictions have received increasing attention from a range of sectors such as agriculture (Vajda and Hyvärinen, 2020), renewable energy (Lledó et al., 2019) and water resource management (Marcos et al., 2017) not only because of its economic potential benefit but also because they could be an advanced support tool to adapt to weather extremes and climate change (Eccel et al., 2016).

The timescale of seasonal predictions ranges between one month and one year (Luo et al., 2011). The predictability at these forecast horizons is linked with both the lower-boundary ocean/land variability (signal) and the variance related to the dynamics within the atmosphere (noise), and it largely comes from the El Niño-Southern Oscillation (ENSO) as well as other slow-evolving phenomena (Doblas-Reyes et al., 2013). The skill level of seasonal prediction systems is one of the key factors affecting users' resolution to include them in their decision-making workflows. Therefore, a systematic evaluation of the predictive performance of the current state-of-the-art seasonal forecasting systems has been undertaken (in raw and bias-corrected conditions) and is described in the following sections.

2.1.1 Data and methods

The monthly predictions of Essential Climate Variables, ECV (2-m mean, maximum and minimum temperatures and total precipitation), from eight forecasting systems from the Copernicus Climate Change Service (C3S) have been downloaded and analysed at monthly time-scale (https://cds.climate.copernicus.eu/cdsapp#!/search?type=dataset). Their acronyms and characteristics are reported in

Table 1. The acronyms consist of the combination of the name of the forecasting centre and the version number of the system used. In addition to the monthly time-frame, these variables have been also aggregated to seasonal level (i.e., three-month average or sum).

In this work, five different calibration techniques have been applied with the aim of correcting the main types of model's biases. Besides, the quality of the calibrated predictions has been compared to that of the raw predictions in order to assess the impact of each correction. The five calibration methods are the following: 'bias' (which corrects only the mean bias; Torralba et al., 2017), 'evmos' (which applies a variance inflation technique to ensure the correspondence of the variance between the predictions and reference data); Van Schaeybroeck and Vannitsem, 2011), 'mse_min' (which corrects the mean bias, the overall forecast variance and the ensemble spread by minimising a constrained mean-squared error; Doblas-Reyes et al., 2005), 'crps_min' (which corrects the mean bias, the overall forecast variance and the ensemble spread while minimising the continuous ranked probability score; Van Schaeybroeck and Vannitsem, 2015) and 'rpc_based' (which adjusts the forecast variance ensuring that the ratio of predictable components is equal to one; Eade et al. 2014). All the calibration methods are implemented in the CSTools R-package (Pérez-Zanón et al., 2022) and have the option to be applied in leave-one-out cross-validation mode (i.e., without using information of the time step that is being calibrated, as it would be done in a real-time context).

Table 1: The eight forecasting systems in the Copernicus Climate Change Service (C3S) and their original configurations. The acronym consists of the combination of the name of the forecasting centre and the version number of the forecasting system used. The verification has been conducted at the coarsest grid (i.e., the JMA-2 resolution at 2.5˚) for the common period, 1993-2016, for all the twelve start dates (i.e., the first day of each month). All the available ensemble members of each individual system have been used. Please note that only the years from 1994 to 2016 have been used for the predictions initialised in January because of the lack of January's predictions in 1993 for UKMO-600.

Acronym	Forecasting system name	Forecasting centre	Hindcast ensemble size	Hindcast period	Horizontal resolution	Reference
ECMWF-5	SEAS5	ECMWF	25	1981-2016	1° × 1°	Johnson et al. (2019)
MeteoFrance-7	System 7	Météo-France	25	1993-2018	1° × 1°	Voldoire et al. (2019)
JMA-2	CPS2	JMA	10	1993-2016	2.5° × 2.5°	Takaya et al. (2018)
ECCC-2	CanCM4i	ECCC	10	1993-2020	1° × 1°	Merryfield et al. (2013)
NCEP-2	CFSv2	NCEP	20	1993-2016	1° × 1°	Saha et al. (2014)
UKMO-600	GloSea6	UK MetOffice	28	1993-2016	1° × 1°	Williams et al. (2018)
DWD-21	GCFS2.1	DWD	30	1993-2019	1° × 1°	Stevens et al. (2013)
CMCC-35	SPS3.5	CMCC	40	1993-2016	1° × 1°	Gualdi et al. (2020)

Due to the varying configurations of the systems included in

Table 1, the verification process over the Southern African Development Community (SADC) has followed these steps:

1. The horizontal resolutions for all the predicted and reference data sets have been upscaled to the coarsest grid (i.e., the JMA-2 at 2.5°), whenever the resolution was finer, with a conservative interpolation method.
2. Although some systems have more years, the common period 1993-2016 has been used for the verification (please note that the predictions initialised in January take into account the years from 1994 onwards because there is no prediction for January of the year 1993 for the system UKMO-600).
3. All the available ensemble members have been used for each individual system.
4. The minimum common number of forecast months for all the systems is six although there are seven months for the ECMWF-5 prediction. Consequently, the lead times for all the twelve start dates (i.e., the first days of each month) range from 0-3 and 0-5 for the seasonal and monthly predictions, respectively.
5. Lastly, the verification metrics that have been systematically computed over each grid point include the mean Bias, the Mean Absolute Error (MAE), the Anomaly Correlation Coefficient (ACC; Wilks, 2011), the Brier Skill Scores (10 and 90) and the fair Ranked Probability Skill Score (fRPSS; Wilks, 2011; Ferro, 2014) based on the tercile equiprobable categories. Besides, the same two-sided t-test and the Random Walk test have been applied to obtain the statistical significance of ACC and fRPSS. In all cases, the reference data set used is the ECMWF ERA5 reanalysis (Hersbach et al., 2020) and the benchmark comparison for all the skill scores, the observed climatology.

2.1.2 Results

The verification process has generated thousands of figures. All of them are available to the consortium partners through a R Shiny App (https://earth.bsc.es/shiny/FOCUS-Africa/). Thus, this section will only offer the analysis of one season and metric (mainly) as an example: JFM fRPSS for 2-m air temperature for the start date of January (0101, hereafter). Hence, the forecast quality measured with the fRPSS for the tercile categories of the seasonal (3-month average) predictions for this season and start date is described in the next section. Whenever the fRPSS is positive, the user is recommended to use the prediction rather than the historical climatology. Conversely, the latter should be used if negative fRPSS is obtained.

The quality of probabilistic forecasts of the seasonal mean of JFM 2-m air temperature for the start date 0101 (first of January) measured with the fRPSS shows varying performances that depend on many factors such as the forecasting system considered and the calibration method applied (as shown in Figure 1). In general, the ECMWF-5 and MeteoFrance-7, regardless of the calibration method applied, show higher skills over the central and central-north areas of the SADC than the other systems. As for the effect of the calibration method applied, the skills for the mse_min and crps_min calibration modes are more similar and usually better when compared to the other methods. Additionally, it is worth noting that for this metric and lead time, almost all the raw predictions outperform the historical climatology over the region.

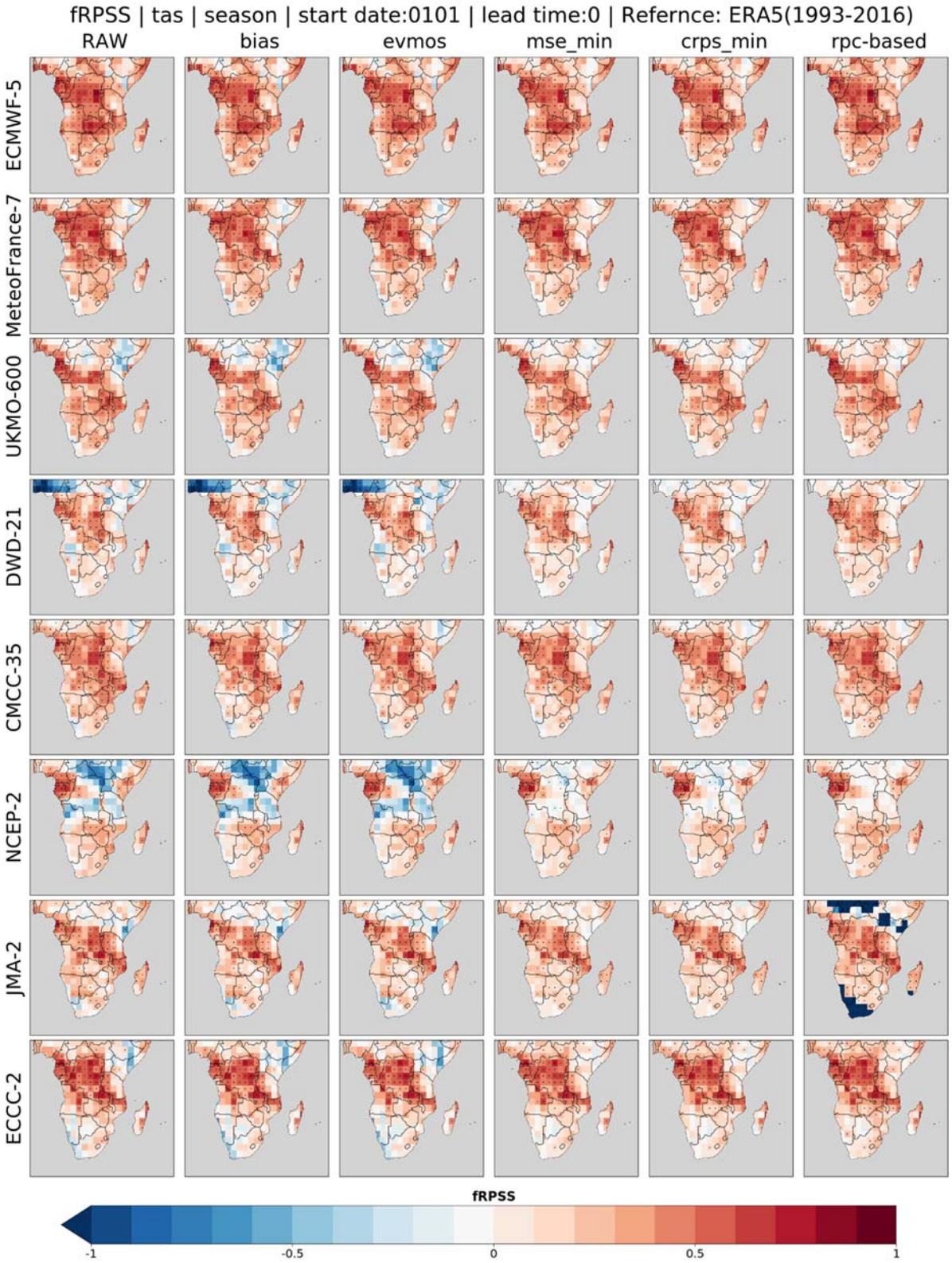


Figure 1: Fair Ranked Probability Skill Scores of the raw and bias-adjusted (column-wise) seasonal predictions of 2-m temperature for the eight C3S forecasting systems (row-wise) for the start date 0101 (first of January) and lead time zero (i.e., JFM mean) for the period from 1994-2016 (the reference data set is the ERA5 reanalysis).

Regarding which approach works better for each grid point, the raw predictions from the eight systems (the 'RAW' column in Figure 1) are compared in Figure 2, where only the system with the highest fRPSS is shown over each grid point of the map. In this case, the ECMWF-5 outperforms the other systems mainly in the southern half of the SADC. Conversely, in the northern regions, the MeteoFrance-7, ECCC-2 and UKMO-600 show a scattered pattern of improvement.

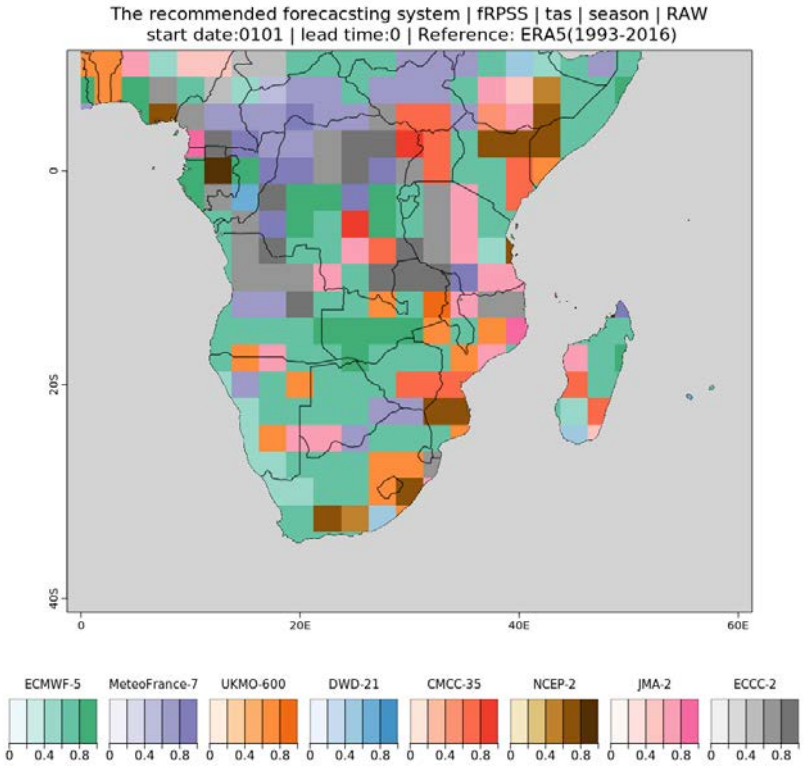


Figure 2: The forecasting system with the highest Fair Ranked Probability Skill Scores over each grid point for the raw seasonal prediction of 2-m temperature for the start date 0101 and lead time zero (i.e., JFM mean) for the period from 1994-2016 (the reference data set is the ERA5 reanalysis).

Regarding the effect of applying bias adjusting methods on the skill metric in each system, Figure 3 shows the 'recommended' method for the ECMWF-5 and JMA-2 systems as examples (from the first and seventh rows in Figure 1). In this case, only the 'rpc-based' method is able to outperform the raw predictions in some areas of both systems. Note that the influence of the application of the methods varies with many factors such as the system, season, etc.

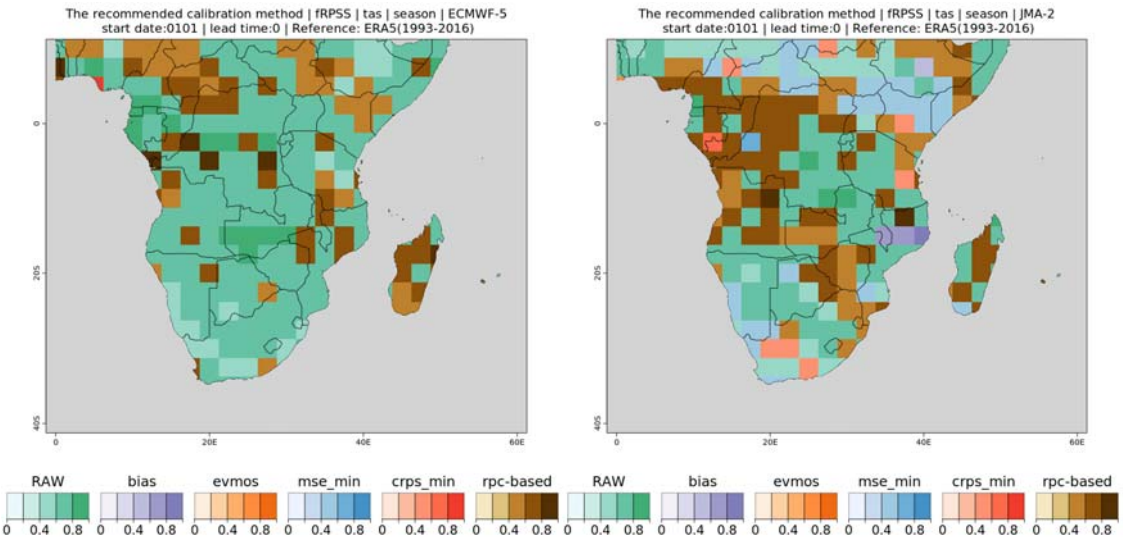


Figure 3: The bias adjusting method with the highest Fair Ranked Probability Skill Scores (fRPSS) over each grid point for the (left) ECMWF-5 and (right) JMA-2 seasonal predictions of 2-m temperature for the start date 0101 and lead time zero (i.e., JFM mean) for the period from 1994-2016 (the reference data set is the ERA5 reanalysis).

One of the advantages of having multiple forecasts available for the users is that even if one region shows poor performance in one system, it might still show good results with another one. For example, the aforementioned 'better' performers (ECMWF-5) have limited/insignificant skills over Tanzania where, conversely, JMA-2 and ECCC-2 display higher and significant skills (see the corresponding rows in Figure 1 and/or Tanzania in Figure 2). In terms of the selection of the calibration methods, its influences on the skill metric of interest could vary with not only the method itself but also the systems applied. For instance, the spatial pattern of the fRPSS shown in Figure 1 remains almost the same for the systems ECMWF-5, MeteoFrance-7 and CMCC-35. However, DWD-21, NCEP-2 and JMA-2 show at least two patterns: one for the 'bias' and 'evmos' modes and another for the 'mse_min' and 'crps_min' (see the corresponding columns). As such, for each grid point, one recommended strategy' (i.e system-method pair) could be selected from the 48 maps analogous to the ones displayed in Figure 1 (i.e., eight forecasting systems + raw plus five calibration modes) based on the statistic of interest for each start date, lead time and variable.

To better visualise the 'recommended strategy' over the SADC, Figure 1 can be summarised into Figure 4 where there is only one strategy with the highest skill metric (e.g., fRPSS in this case) shown over each grid point. As for the legend, each forecasting system uses one 'leading' colour with different tones: green for ECMWF-5, purple for MeteoFrance-7, orange for UKMO-600, blue for DWD-21, red for CMCC-35, brown for NCEP-2, pink for JMA-2 and grey for ECCC-2. Furthermore, the six gradients of each main colour represent the six post processing approaches: 'RAW', 'bias', 'evmos', 'mse_min', 'crps_min' and 'rpc-based' (from darker to lighter shades).

For JFM 2-m air temperature prediction, the ECMWF-5 raw prediction shows the highest fRPSS over more grid cells in the southern and north-eastern SADC. The 'RAW' and 'rpc-based' predictions from the MeteoFrance-7 and ECCC-2 systems, on the other hand, can be found over the northwestern and central-northern regions, with the remaining strategies showing only in scattered grid points.

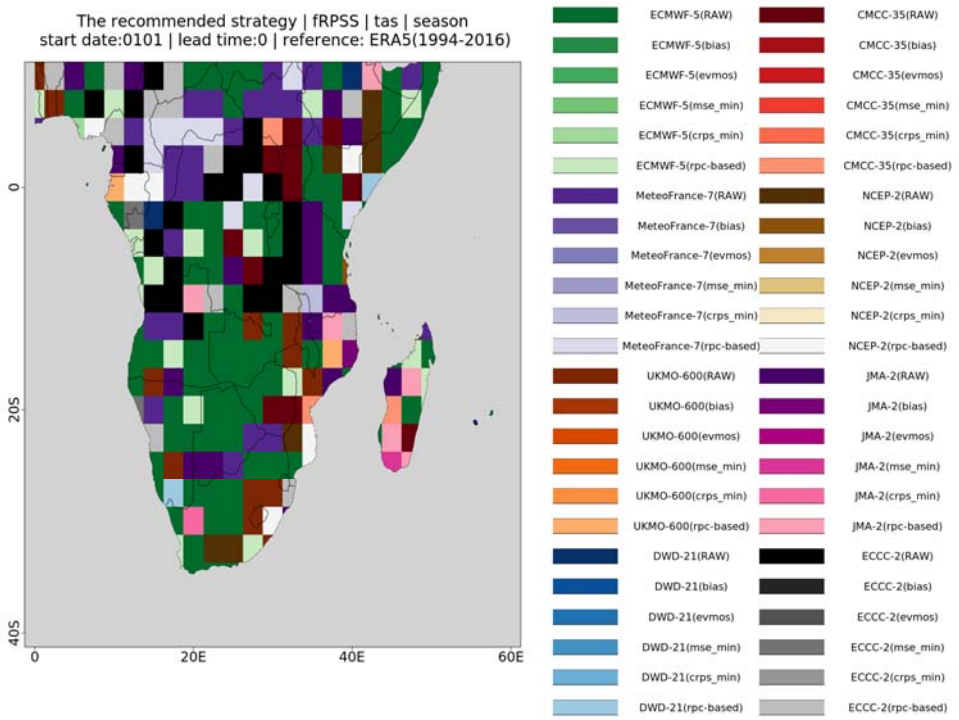


Figure 4: The recommended strategy (i.e., the system-method pair with the highest fRPSS) for the seasonal predictions of 2-m temperature for the start date 0101 with the lead time zero over the 1994-2016 period. The reference data set used is the ECMWF ERA5 reanalysis. In the legend, each forecasting system uses one main colour: green for ECMWF-5, purple for MeteoFrance-7, orange for UKMO-600, blue for DWD-21, red for CMCC-35, brown for NCEP-2, pink for JMA-2 and grey for ECCC-2. The six gradients of each main colour represent the six calibration modes: 'RAW', 'bias', 'evmos', 'mse_min', 'crps_min' and 'rpc-based' from dark to light.

In Figure 4 the spatial distribution for one lead time of one skill metric can be observed. However, the recommended 'strategy' may vary with lead times and/or metrics. Therefore, Figure 4 could be further summarised in a pie chart by computing the percentages of the grid points for the 'recommended performers' over the entire SADC. Although the spatial pattern would not be available in the pie chart, the variability of the percentages for all the five skill metrics at different lead times could still be obtained. Therefore, Figure 5 depicts the shares of the recommended strategies for all metrics (column-wise, ACC, BrierSS10, BrierSS90, MAE and fRPSS from left to right) and four lead times from zero to three (row wise) for the same predictions of JFM 2-m air temperature.

For all skill metrics at the lead time zero (the top row), when looking at the shares of each forecasting system (regardless of the calibration method), the ECMWF-5 is the predominant performer over around a quarter (for BrierSS90) to a half (for ACC) of the whole SADC. Moreover, its dominance basically remained throughout all the lead times for all skill metrics except for BrierSS10 whose percentage decreased to about ten at lead time three. Thus, a good level of skills (except for BrierSS10) can be expected when using raw and bias-adjusted ECMWF-5 predictions for the JFM 2-m air temperature predictions (start date 0101) over widespread areas of the SADC.

In fact, the use and selection of the calibration methods mainly depend on the skill metrics of interest. For example, applying a calibration method barely increased the ACC because almost only the darkest colors could be seen for each system (see the first column). However, for BrierSS10, BrierSS90 and fRPSS, some lighter colors begin to show in the charts. These increases in the percentages of the bias

adjusted predictions hint at the added value of the calibration method applied. Regarding the MAE, it shows the widest spectrum of each main color especially at lead time three. Hence, in this case it is more advisable to apply a bias correction method. Finally, considering the variability of performance of the forecasting systems with lead times, it is worth noting here that NCEP-2 (ECCC-2) shows a marked improvement with the lead time, increasing from 5-6% to around 20% for BrierSS90 (BrierSS10).

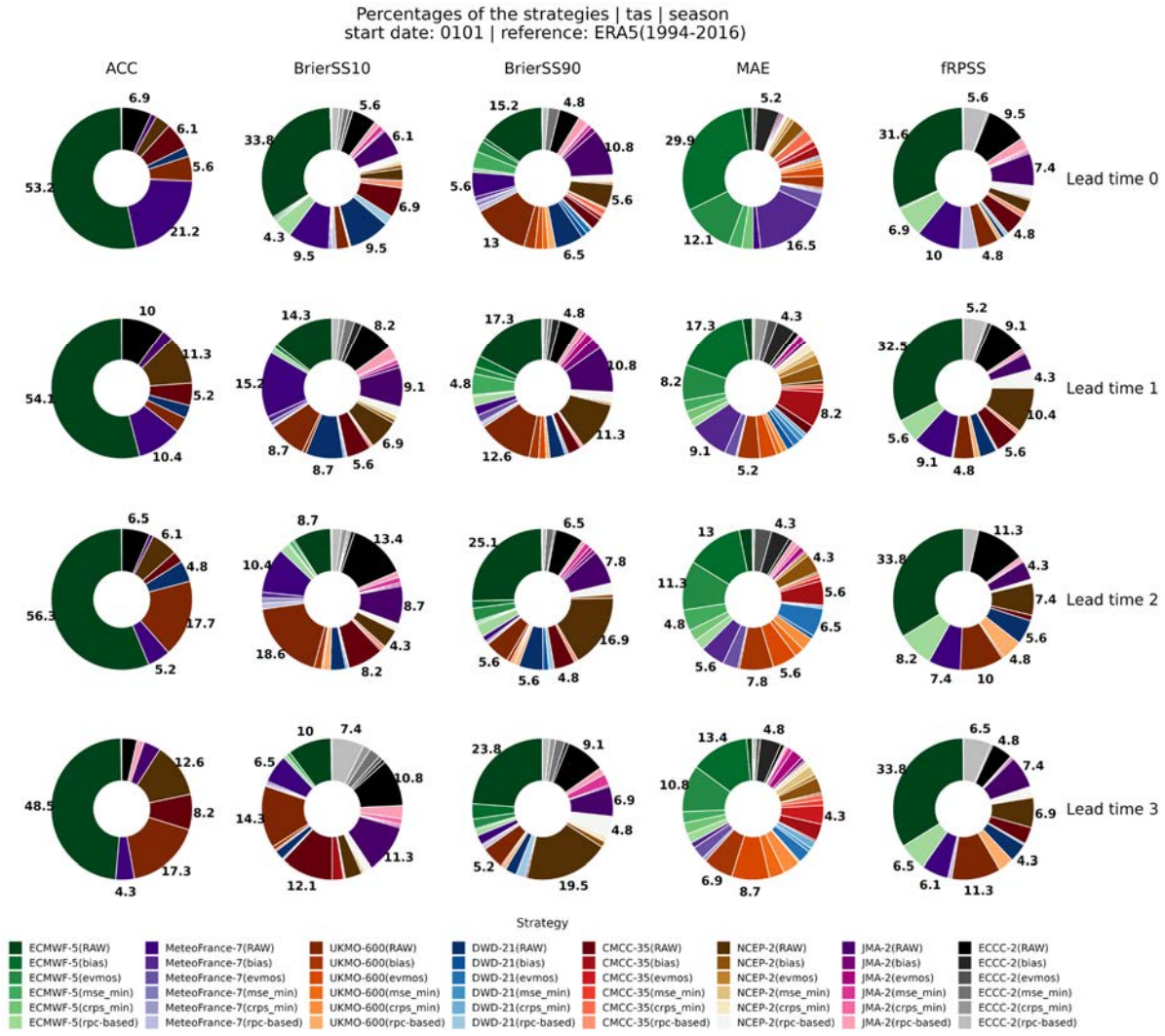


Figure 5: Percentages of the grid points for the recommended strategies (see the legend) over the entire SADC for the prediction of JFM 2-m air temperature for the start date 0101 based on the corresponding skill metrics (column-wise, ACC, BrierSS10, BrierSS90, MAE and fRPSS from left to right) for the period from 1994-2016. The lead times range from zero to three by rows (top to bottom). In the legend, each forecasting system uses one main colour: green for ECMWF-5, purple for MeteoFrance-7, orange for UKMO-600, blue for DWD-21, red for CMCC-35, brown for NCEP-2, pink for JMA-2 and grey for ECCC-2. The six gradients of each main colour represent the six calibration modes: 'RAW', 'bias', 'evmos', 'mse_min', 'crps_min' and 'rpc-based' from dark to light. The numbers on the labels of the 'pies' represent the percentages of the corresponding colours (strategies). To make the figure tidy, the labels of the percentages with fewer than ten grid points were removed.

2.2 Bias correction of extremes in seasonal forecasts

The raw output of regional and global climate models can be immediately used, without any further processing, to analyse long-term tendencies (e.g. decadal time scale and beyond) and shorter term anomalies (e.g. seasonal to decadal time scale). However, using climate model outputs for impact studies in any sector of human activities (agriculture, energy, health, safety, infrastructures etc.) requires a preliminary processing to remove any systematic error coming from the model. Very often, impact models describe processes of a different nature which depend on specific critical threshold linked to well defined numerical values of environmental parameters. Examples include the modelling of biological systems, including agronomic models or models for the diffusion of pests and diseases, or the modelling of landslides and flash floods.

Bias correction algorithms are specifically designed to remove systematic model errors and provide an effective interface between climate and impact models. A systematic error is a known artefact of a numerical climate model, which is produced by some undefined mechanism related to the complex nature of the model itself, such as the interaction between the different components of a climate model, or the use of approximations to model the physics of fundamental processes. Systematic errors entail the unrealistic constants drifts of climate variables (e.g. temperature), which generally requires a simple removal of the cumulated difference (*delta)* between model and observations, or the misrepresentation of the statistical distribution of climate variables, that can be cured by remapping the model output to some reference statistical distribution. More sophisticated bias correction algorithms may impose other important properties to the model output such as the constraint to specific spatial patterns (Maraun and Widmann, 2018). Another class of bias correction methods ensure the correct interdependence of related climate variables, such as temperature and precipitation, in order to retain the physical coherence that gets lost if the variables are treated separately (e.g Cannon et al., 2016 and Vrac, 2018).

The specific challenge of this work is to design a bias correction method that provides an extrapolation of extreme temperature and rainfall events, usually localised in their geographical extent. Indeed, one of the limitations of bias correction methods based on quantile mapping is that they are not good at dealing with the tails of Cumulative Distribution Functions (CDFs) (White and Toumi, 2013). Therefore, the proposed approach consists of conceptually extending one of the classic quantile mapping methods by improving the description of the tail of the distribution. The methodology developed for this study is based on fitting a parametric transformation to the quantile-quantile relation of observed and modelled values over the core of the distribution. For the tail of the distribution, the quantile-quantile relation of the modelled data is remapped to the appropriate generalised extreme value (GEV) distribution.

2.2.1 Data and methods

The methodology proposed in this work is used to perform the bias correction of extreme 2-m temperature and total precipitation for three forecasting systems: ECMWF, Météo-France and DWD. The meaning of their acronyms, as well as their characteristics, are listed in

Table 1. The analysis has been performed over the Southern African Development Community (SADC) during the 1993-2020 period.

The proposed technique is the combination of two approaches: quantile mapping (QM) and a generalized extreme value distribution (GEV) fitting. QM is applied to the central part of the statistical distribution whereas the GEV is applied to the tails of the distribution (i.e., for the extremes). Since there is no special library in Python that does this work, the bias correction algorithm has been implemented in a new ad hoc class.

Quantile mapping has been widely used and is among the most important and popular bias correction methods. It assumes that the cumulative distribution function (CDF) of a variable in the forecast and observation time series does not change in the future period (Y. Tong et al., 2020). Given a variable *x*, QM minimises the discrepancy between the CDF of the model data and that of the reference data, over a certain calibration period. In practice, in a quantile mapping algorithm the model output *x* is mapped to an observation output *y* by means of a transform function *h*, in such a way that their two CDFs are equivalent (C. Piani et al., 2010):

$$y = h(x) \rightarrow CDF_y(y) = CDF_x(x)$$
$$y = {CDF_y}^{-1}(CDF_x(x))$$

Where CDF^{-1} is the inverse function of the CDF. Generalised extreme value is a family of continuous probability distributions often used for modelling extreme events. The GEV has three parameters: location, μ, scale, σ, and one shape, ξ, parameter. Depending on the value of the latter, the GEV distribution is classified as Gumbel (shape parameter equal to 0), Fréchet (shape parameter greater than 0), and Weibull (shape parameter lower than 0). The cumulative distribution function (CDF) of the GEV distribution is the following:

$$CDF(x;\mu;\sigma;\xi) = exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

In the case of total precipitation, the definition of extremes corresponds to the values above the 95th percentile among the time series data. Considering temperature, also the lower extremes are taken into account, i.e., the values below the 5th percentile. Hence, precipitation data is bias corrected using the classic quantile mapping below the 95th percentile, and the GEV fitting above that threshold. Similarly, for temperature the quantile mapping is applied to the data between the 5th and 95th percentile, while GEV is fitted to the tails of the statistical distribution.

In Python, there are a few libraries for performing bias correction. For example, the *bias_correction* module, released in August 2021 (https://pypi.org/project/bias-correction/), consists of functions to remove biases across datasets. Implemented methods include quantile mapping, modified quantile mapping and scaled distribution mapping (Gamma and Normal Corrections). This module takes as input three datasets: an observation (reference) dataset, a dataset containing model simulations during the same period as the observation data, and the model projection for a future period. The *bias_correction* module gives as output the bias-corrected dataset for the future period. The downside of this tool is that at every usage the reference dataset and the model simulations during the same

period are required. In this way, the training phase has to be repeated every time, increasing the computational resources required. For this reason, an ad hoc Python class has been created, which separates the preparation of the bias correction model over a series of years (usually about 30) from the operational use. The parameters of the bias correction models are saved and can be used in the future applied to any period (without the need to recalculate them). The procedure consists of two phases:

- In the first phase, which can be called the 'training' phase, the parameters of the two statistical distributions are computed (the one corresponding to be bias corrected, and the reference one). In our case, the data to be corrected are the seasonal forecasts provided by ECMWF, Météo-France and DWD, while as a reference, the ERA5 reanalysis is used. Since the applied method is different depending on whether the data is extreme or not, there will be two types of parameters: for the central part of the distribution the parameters are the CDF values corresponding to 20 quantiles (whose number can be changed by specifying it in the library). Regarding the distribution tails, the parameters will be the 3 characteristics of the GEV: location μ, scale σ, and shape ξ.

- In the second phase, namely the application of the bias correction technique, the parameters obtained during the training phase are used to correct the seasonal forecast data by applying the inverse function of the CDF.

2.2.2 Results

In this section, the raw predictions for three forecast centers (ECMWF, Météo-France and DWD) have been evaluated and compared against the different bias-corrected predictions (simple QM and the new method).

The procedure described in the previous section is applied for each month (e.g., for all Januaries from 1994 to 2020). Here, the months of January, February and March are shown as an example. Furthermore, the results are different depending on the 'lead time'. Indeed, in seasonal forecasts, predictions of the same month are provided at different starting months and these outputs are identified by a different 'lead time'. For example, the forecast of January 2020 provided by the seasonal forecast initialised in December 2019: in this case, the lead time would be 1. Instead, the forecast of January 2020 initialised in July 2019 is identified by lead time 6. Since forecasts corresponding to different lead times can have diverse performances in terms of accuracy, it has been decided to consider them independently. Hence, for each month 6 results will be produced, corresponding to the 6 lead times.

A first evaluation of the performance of the proposed bias correction method consists in comparing the extremes of the seasonal forecast corresponding to two percentiles, 97^{th} and 99^{th}, with the respective values of the reanalysis ERA5. For this purpose, the difference between the seasonal forecast and the reference is shown in three situations: before applying the bias correction (i.e., using raw data), after applying the classic quantile mapping approach, and after using the proposed method (i.e., the combination of quantile mapping and GEV). To perform all these operations the reanalysis

data had to be previously re-gridded onto the same coordinates of the seasonal forecast, which have a resolution of 1°, using a nearest neighbors' interpolation. The results obtained for 2-m temperature with the three models (SEAS5, System 7 and GCFS2.1) for January and lead time 1 are shown in Figure 6,Figure **7** and Figure **8** as an example. In these figures, the left-hand side panel refers to the difference calculated with the raw data, while the centre and right-hand side panels refer to data bias corrected by quantile mapping and by our proposed method, respectively. For the sake of brevity, only the results obtained on the 99th percentile will be shown hereafter.

As it can be seen from the left-hand side panels, raw seasonal forecasts underestimate the temperature extremes corresponding to the 99th percentile in almost the entire study area. This underestimation is more pronounced for SEAS5, while System 7 and GCFS2.1 behave differently in some areas. The application of quantile mapping produces a clear decrease of the discrepancy, even if a negative bias remains in the southwestern part of the African continent. The use of GEV fitting for the tails produces a further improvement on the results, as it can be seen from the minor differences between bias-corrected data and reanalysis. Similar outcomes are also found for the 97th percentile and for the other months and lead times.

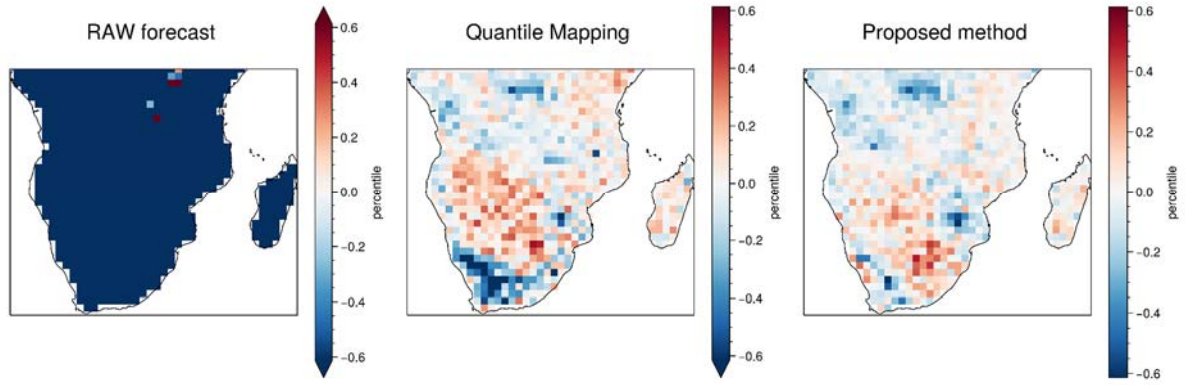


Figure 6: Difference between the 2-m mean temperature 99th percentile value for seasonal forecast SEAS5 (ECMWF) and ERA5 reanalysis. Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast. Maps refer to January and lead time 1.

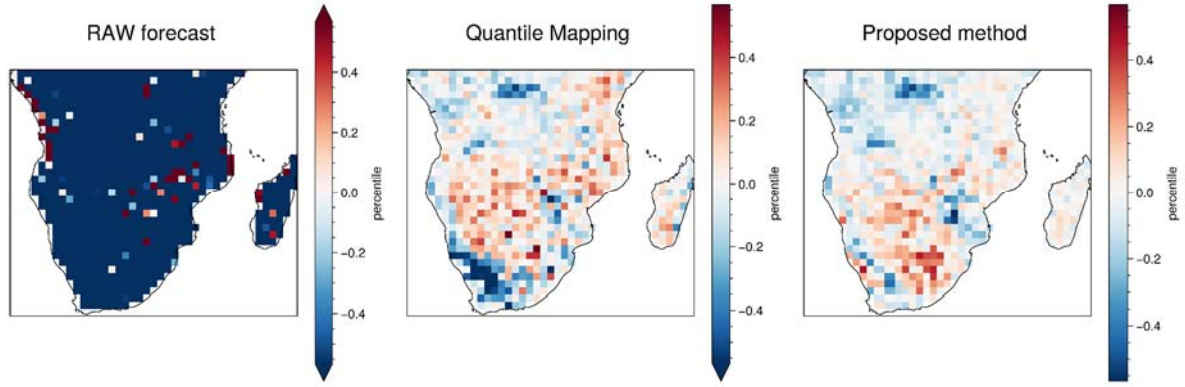


Figure 7: Difference between the 2-m mean temperature 99th percentile value for seasonal forecast System 7 (Météo-France) and ERA5 reanalysis. Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast. Maps refer to January and lead time 1.

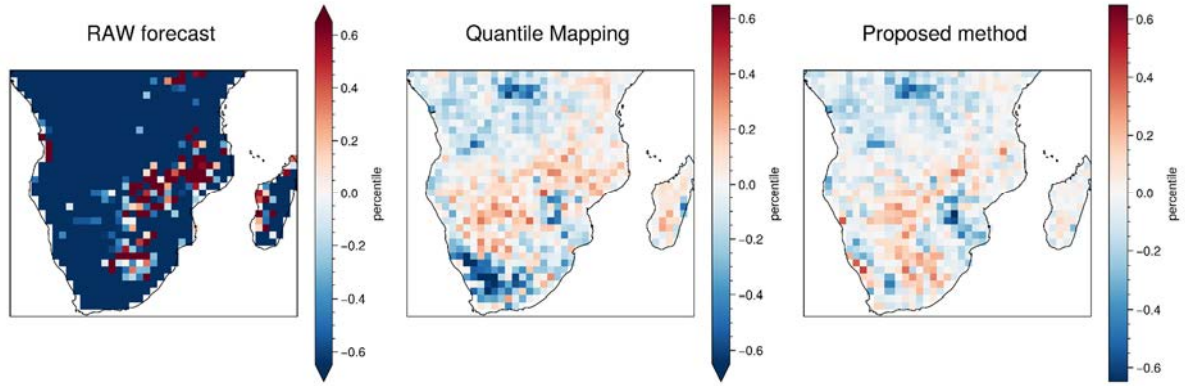


Figure 8: Difference between the 2-m mean temperature 99th percentile value for seasonal forecast GCFS2.1 (DWD) and ERA5 reanalysis. Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast. Maps refer to January and lead time 1.

A good indicator of a model's performance is the Root Mean Squared Error (RMSE), which measures the deviation between predicted and reference data. For both variables, temperature and precipitation, the RMSE between the extremes at the 99th and 97th percentile of the seasonal forecasts and those of ERA5 has been computed. As it has been previously done, the calculation has been performed both for raw and corrected data.

Concerning temperature, the values obtained are summarised in Figure 9Figure 10 and Figure **11**, which show the RMSE related to each month and lead time. Both types of bias correction improve the performance of the predictions from the RMSE point of view, as it can be seen from the lowering of this metric. Additionally, the proposed methodology further lowers the RMSE compared to quantile mapping alone. Raw seasonal forecasts of 2-m temperature have a RMSE which in addition to being larger, tends to vary with lead time. In some cases, like for SEAS5 and GCFS2.1, it decreases with increasing lead time, contrary to what is expected. However, when bias-corrected, the RMSE becomes much smaller and varies slightly, independently of lead time.

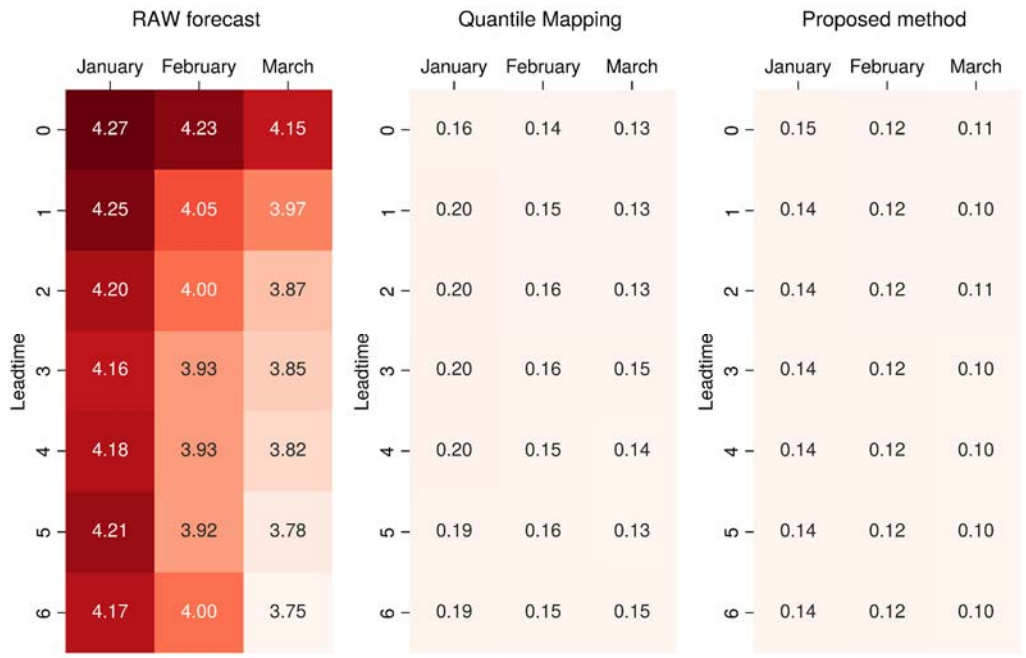


Figure 9: RMSE of the 2-m mean temperature 99th percentile value for seasonal forecast SEAS5 (ECMWF) compared to ERA5 reanalysis. Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast.

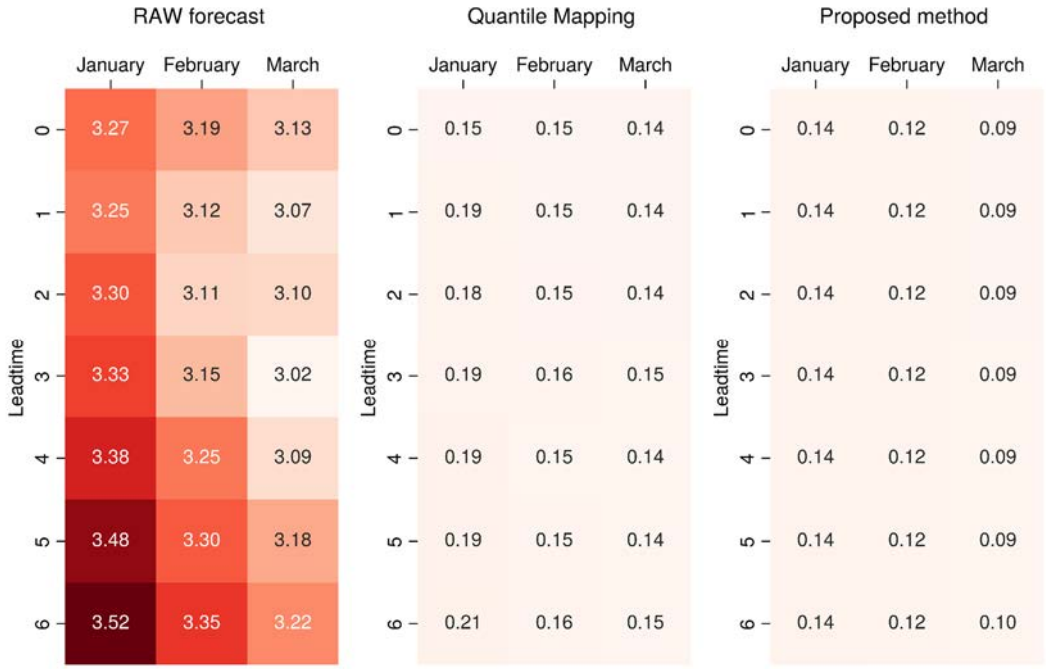


Figure 10: RMSE of the 2-m mean temperature 99th percentile value for seasonal forecast System 7 (Météo-France) compared to ERA5 reanalysis. Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast.

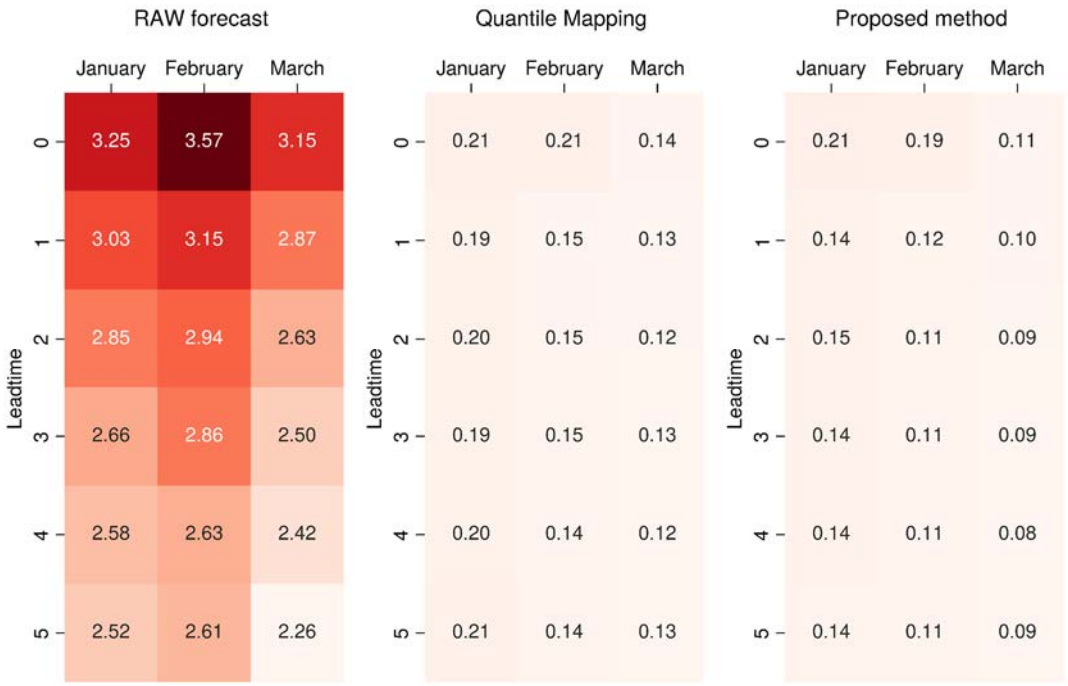


Figure 11: RMSE (K) of the 2-m mean temperature 99th percentile value for seasonal forecast GCFS2.1 (DWD) compared to ERA5 reanalysis. Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast.

This behaviour is also observed in precipitation, where there is a consistent decrease of the RMSE with lead time (Figure 13 and Figure **14**). The application of bias correction produces a decrease in RMSE of one order of magnitude and, sometimes, for System 7, Météo-France (Figure 8) and GCFS2.1, DWD (Figure 14), the decrease is of two orders of magnitude.

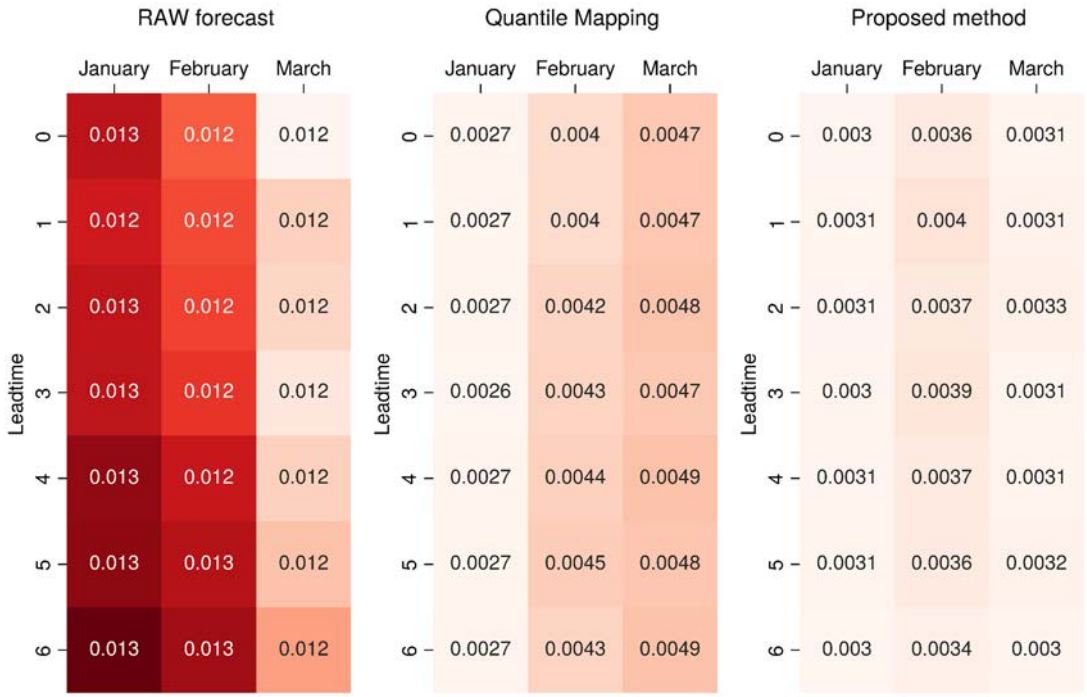


Figure 12: RMSE (m) of the total precipitation 99th percentile value for seasonal forecast SEAS5 (ECMWF) compared to ERA5 reanalysis. Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast.

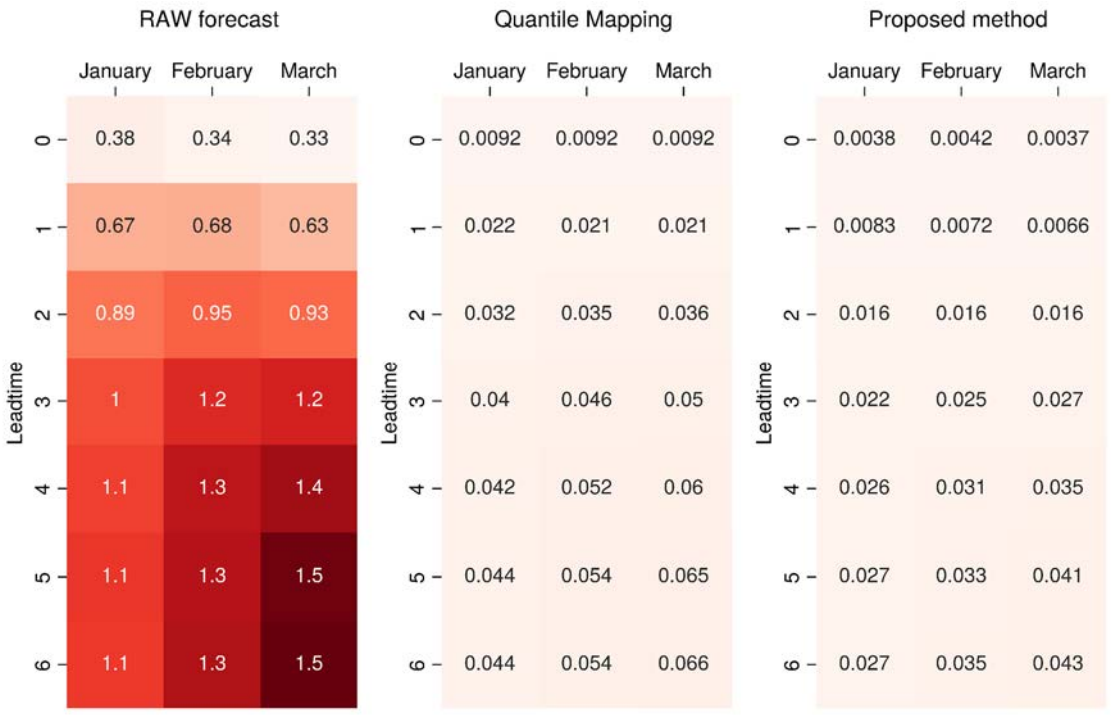


Figure 13: RMSE (m) of the total precipitation 99th percentile value for seasonal forecast System 7 (Météo-France) compared to ERA5 reanalysis. Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias-correction method (right-hand side panel) seasonal forecast.

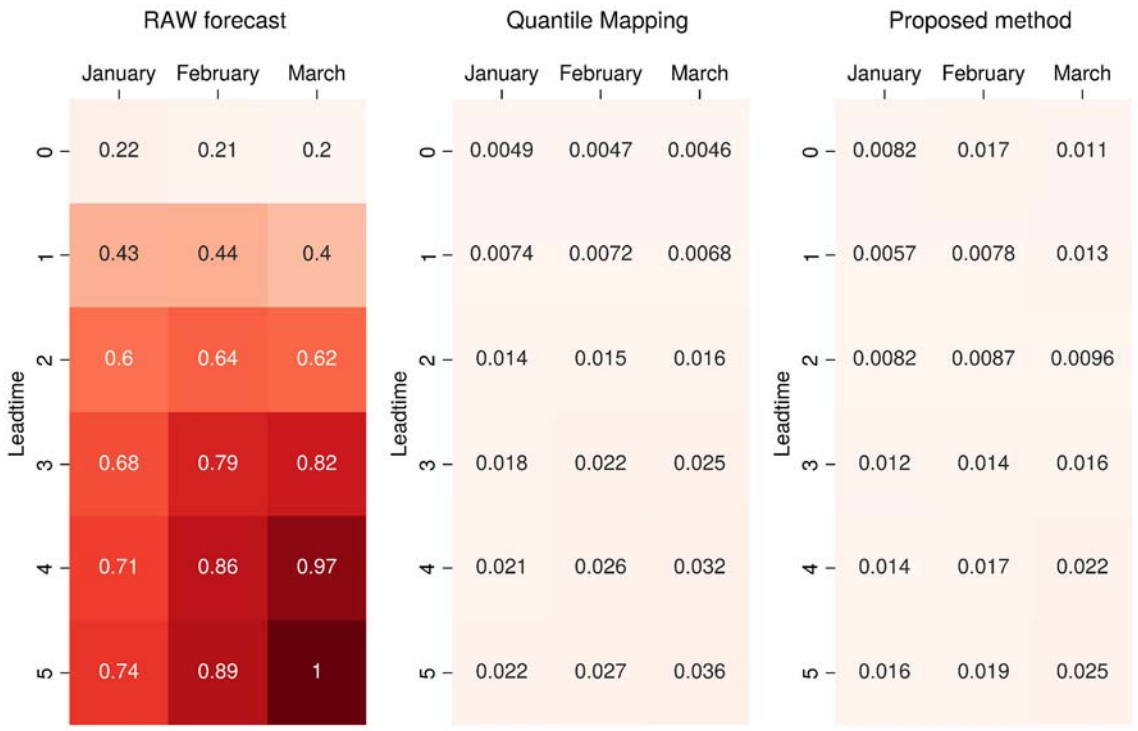


Figure 14: RMSE (m) of the total precipitation 99th percentile value for seasonal forecast GCFS2.1 (DWD) compared to ERA5 reanalysis. Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast.

It is also interesting to see how the percentage of ensemble members which identify extremes varies, before and after applying bias correction. Indeed, seasonal forecasts consist of a certain number of ensemble members, i.e., independent realizations of the forecast. In the case of SEAS5, there are 25 members for hindcasts (period from 1993 to 2016) and 51 for real-time forecasts (since 2017). System 7 also has 25 members, while GCFS2.1 has 30. Each member is an independent forecast with different initial conditions and the spread of the ensemble members provides information on the confidence of the prediction.

First, the extreme events have been identified according to the ERA5 reanalysis. This analysis has been performed for every grid point, considering 'extreme' any value above the threshold (97th or 99th percentile). Then, for each model ensemble, the number of ensemble members overcoming the threshold at least once in a time-window centered on the date of the event under analysis has been computed. As these are long-term forecasts, it makes no sense to look at the exact date on which the extreme event occurred, because the seasonal forecast is unlikely to identify the phenomenon on that specific day. For this reason, a 14-day window centered on each date of the seasonal forecast dataset has been considered (the previous and the following weeks). In this time window, the maximum value of the variable has been identified and compared with the threshold value. If the maximum value is higher than the quantile, the threshold is exceeded at least once, and, according to the proposed definition, that ensemble member identifies an extreme event.

Then, for each cell and timestep, the number of ensemble members for which the threshold was exceeded has been counted. Afterwards this, the sum is divided by the total number of ensemble members, obtaining, for each event, the percentage of members that have correctly detected the

extremes. Finally, for each cell the average percentage has been computed. This calculation has been done for every month and lead time. In Figure 15Figure 16Figure 17, the maps containing the percentage for each cell, referred to January, February and March predicted with lead time 1 are shown as an example. Only raw data and data corrected with the method proposed are shown (quantile mapping alone has produced very similar results to the proposed method). In general, the proposed bias correction improves the ability of ensemble members to predict extreme events.

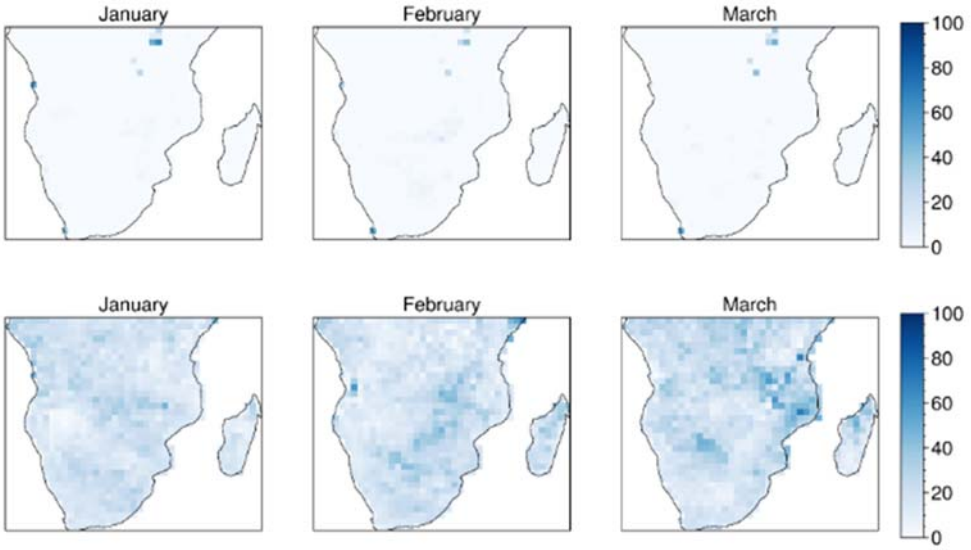


Figure 15: Spatial representation of the ensemble members percentage correctly predicting 2-m mean temperature over the 99th percentile value for raw (first row) and bias-corrected (second row) seasonal forecast SEAS5 (ECMWF). Results are shown for the three months (row-wise) and lead time 1.

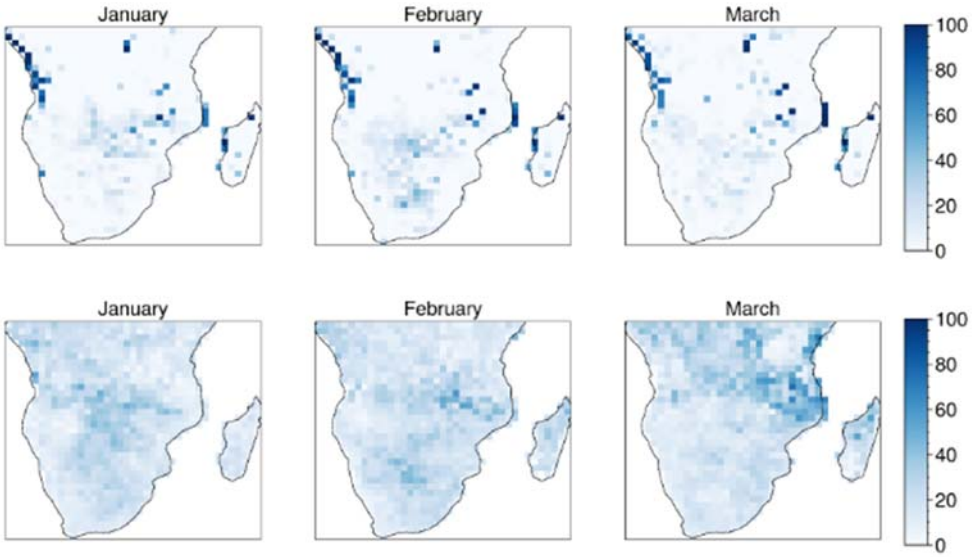


Figure 16: Spatial representation of the ensemble members percentage correctly predicting 2-m mean temperature over the 99th percentile value for raw (first row) and bias-corrected (second row) seasonal forecast System 7 (Météo-France). Results are shown for the three months (row-wise) and lead time 1.

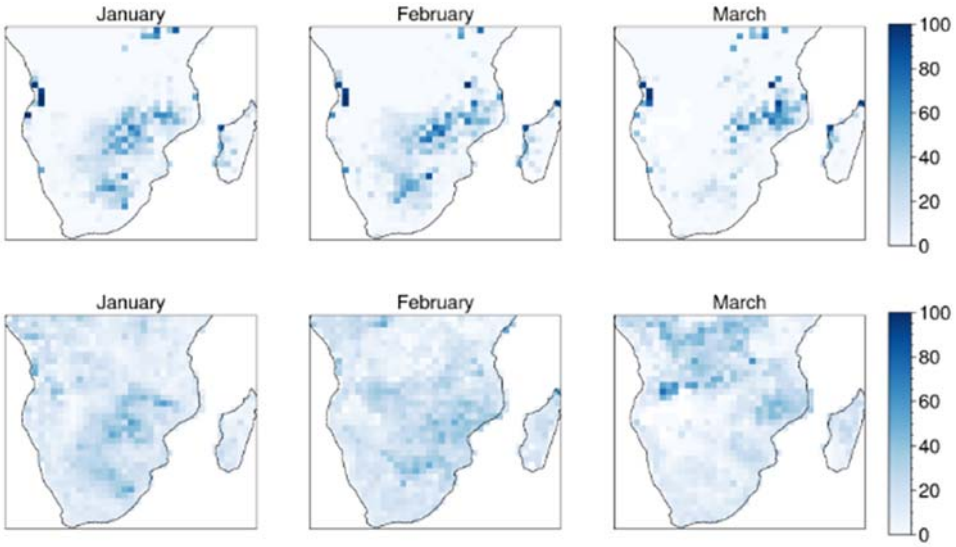


Figure 17: Spatial representation of the ensemble members percentage correctly predicting 2-m mean temperature over the 99th percentile value for raw (first row) and bias-corrected (second row) seasonal forecast GCFS2.1 (DWD). Results are shown for the three months (row-wise) and lead time 1.

The percentages for each month and lead time, averaged over the entire domain, are summarized in the tables corresponding to Figure 18, Figure **19** and Figure **20**. As it can be seen from these tables, the proposed bias correction method produces a net increase in the percentage of members able to correctly predict temperature extremes. The improvement is particularly drastic for SEAS5, whose raw data presents a very low percentage, under 1%. Thanks to the correction, the percentage range increases from 16 to 23% depending on month and lead time. In general, performance is better in March and tends to decrease as lead time increases.

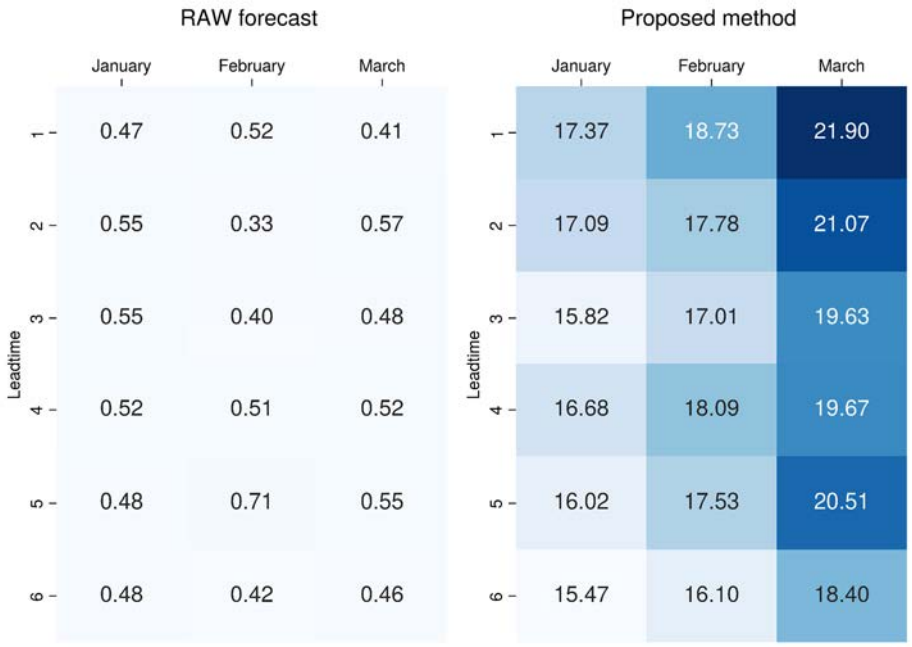


Figure 18: Mean percentage of ensemble members correctly predicting 2-m mean temperature over the 99th percentile value for seasonal forecast SEAS5 (ECMWF). Raw (left-hand side panel), and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast.

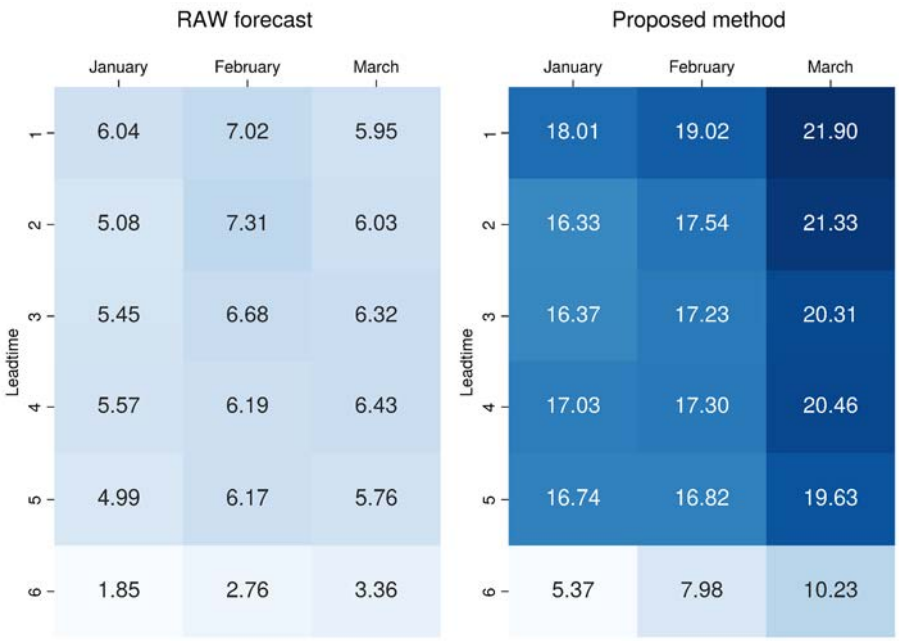


Figure 19: Mean percentage of ensemble members correctly predicting 2-m mean temperature over the 99th percentile value for seasonal forecast System 7 (Météo-France). Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast.

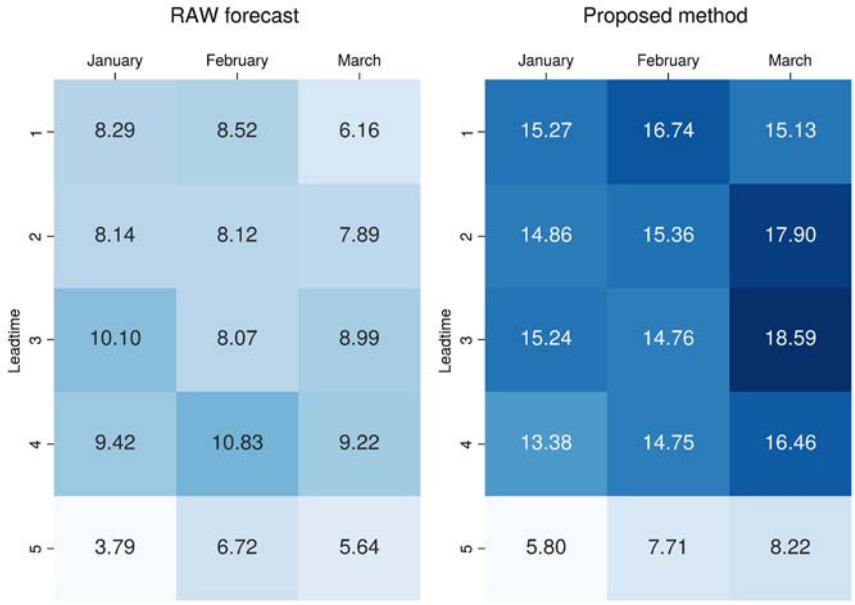


Figure 20: Mean percentage of ensemble members correctly predicting 2-m mean temperature over the 99th percentile value for seasonal forecast GCFS2.1 (DWD). Raw (left-hand side panel), corrected with quantile mapping (centre panel) and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast.

The situation is different for precipitation, in which the correction produces a lowering of the percentages. The result shown in Figure 21 refers to SEAS5, as System 7 and GCFS2.1 have similar behaviours. After the correction there is a drop in the percentage of ensemble members in accordance with the extreme events detected by ERA5. However, it is interesting to note that the percentage of the bias-corrected data does not appear to depend on month or lead time.

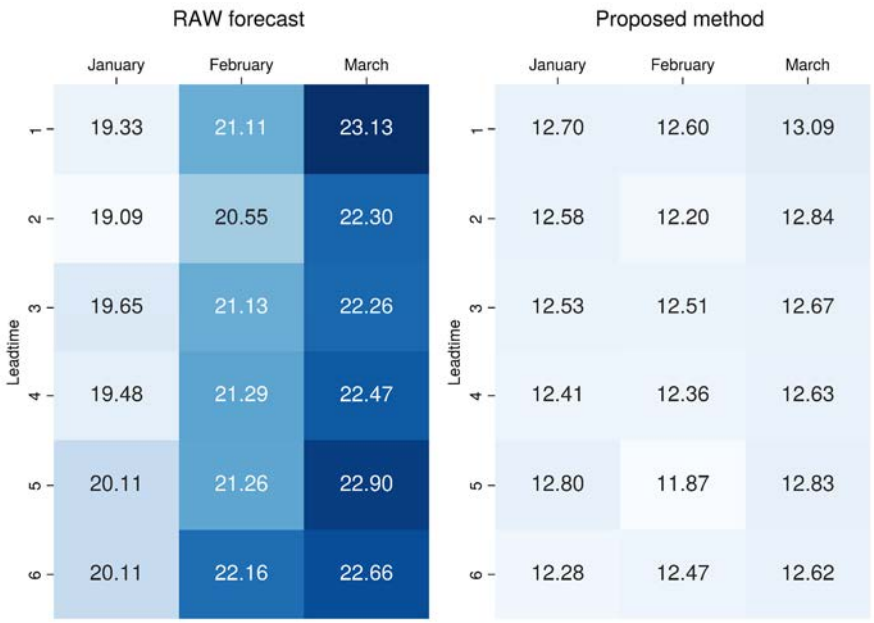


Figure 21: Mean percentage of ensemble members correctly predicting total precipitation over the 99th percentile value for seasonal forecast SEAS5 (ECMWF). Raw (left-hand side panel), and corrected with the proposed bias correction method (right-hand side panel) seasonal forecast.

2.3 Skill of seasonal forecasts for hydrological applications

The objective of this section is to provide an evaluation of the skill of seasonal forecasts over SADC and the Malawi Shire's basin, with a focus on hydrological analyses (and, by implication, water and energy resources). This objective has been translated into the following analysis framework:

- Evaluation is done for hydrological basins (Figure 22) rather than for climate model grid (although there are still some analyses conducted at grid level).
- The primary focus is on forecasting rainfall over the main part of the rainy season, which in the vast majority of SADC region is the DJF season, with the forecast issued at a range of lead times. This focus reflects the primary concern of water resource managers, i.e. the state of water resources at the end of the rainy season (although the last month of DJF, i.e. Feb is technically not the end of that season).
- Evaluation is done for individual models from the Copernicus Climate Change Service (C3S) obtained through the Copernicus Climate Data Store (cds.climate.copernicus.eu). The main goal of this approach is to enable the selection of skillful models/forecasts for the subsequent hydrological simulations (or to build a multi-model ensemble for these simulations).
- In addition to the Copernicus systems, the skill of the Southern African Regional Climate Outlook Forum (SARCOF) forecast has also been evaluated. The SARCOF forecast is based on the so-called consensus approach, whereby different information sources are interpreted by experts towards derivation of a tercile forecast for seasons at a range of lead times. As such, SARCOF forecast cannot be directly linked to hydrological or water resource models, as it does not contain any information other than probabilities of tercile categories for the target season. However,

- SARCOF is an established process, well embedded in institutional and decision-making landscape of the SADC region, and as such it is a 'reference' for any other seasonal forecasting activity.
- SARCOF forecast is issued for lead times that exceed those of the dynamical forecasts from Copernicus and, as such, it may be a source of 'soft' information contributing to decisions related to operations of water resources/energy systems in addition to 'hard' information emerging from the numerical forecasting systems.

- The focus is on the standard probabilistic tercile forecast, i.e. forecast of probabilities of normal, below normal, above normal categories, with the two latter effectively translating into a one-in-three-year drought and one-in-three-year wet conditions.

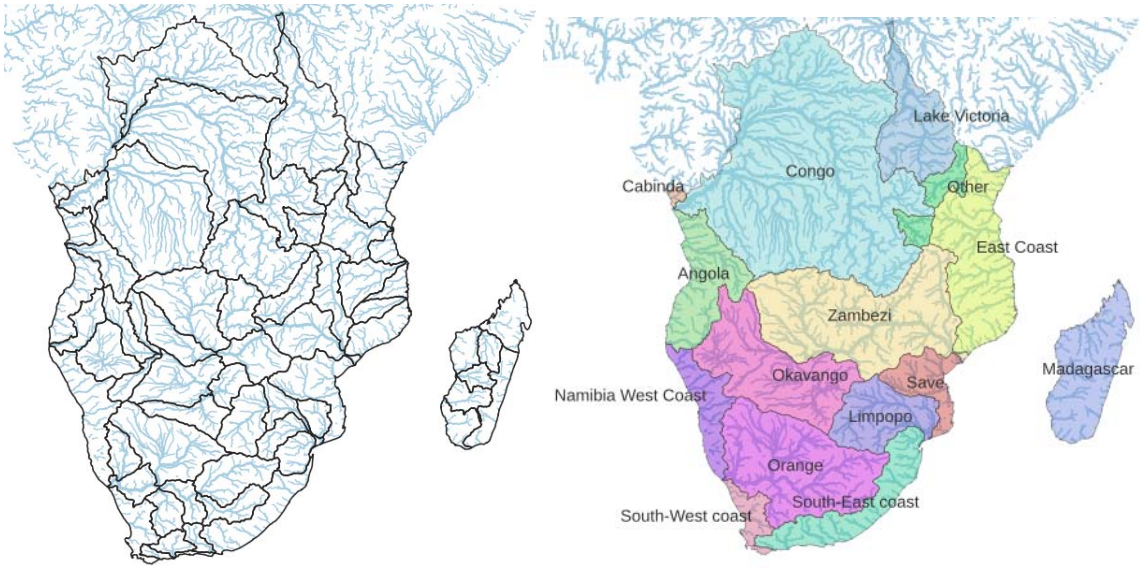


Figure 22: 63 main hydrological basins in the SADC region (left) and mair river basins (right).

2.3.1 Data and methods

2.3.1.1 Dynamical models from the Copernicus Climate Data Store

Currently (September 2022), that resource provides access to forecasts from nine forecasting systems. However, in this subsection only eight have been analysed, due to the complexity of compiling appropriate hindcast data for the GloSea6 system which generates hindcasts on-the-fly. Descriptions of the forecasting systems used are available at the following link, while the summary of available data can be found at this other link. The main characteristics of the data streams for the Copernicus forecasts are as follows:

- Forecasts are issued around the 15th of every month.
- The initialization time of the forecasts varies between systems. Some are initialized on the first day of the month (all members are initialized on the same day/time), others are intialized several times in a month (all member of the ensemble are intialized several times in a month) whereas some have a staggered initialization (each member is initialized on a different day/time). In this regard, CDS consolidates data from all the systems into a uniform

framework, namely - all members for each system are presented as starting on the first day of the forecast month.

- Forecast time series spans 6 months including the month on which the forecast is issued
- Individual models provide between 10 and 51 ensemble members
- Data are available on daily time step with spatial resolution varying between systems and falling between 0.4-2.8 deg (approx. 40-280km)
- Rainfall is one of many available variables.
- Hindcast (or retrospective forecast) data are available for different periods depending on the model, but in principle, all models have hindcasts between 1993 and 2016.

2.3.1.2 SARCOF forecast

The SARCOF forecast is based on the so-called consensus approach, whereby different data sources are interpreted by experts towards derivation of a tercile forecast for seasons at a range of lead times. SARCOF forecast is issued twice a year - in August and in December or January (the so-called SARCOF update). In August, the forecast is generated for the following seasons: OND, NDJ, DJF, JFM. On the other hand, in December/January, forecasts are generated for FMA and MAM. In this work only August forecasts are analysed as they are more relevant from the perspective of this activity. Moreover, data from SARCOF updated forecasts are only available for several years, which is not enough for appropriate skill evaluation. SARCOF forecasts after 2011 are analysed (Figure 23), as only these were 'recovered' from available archived materials.

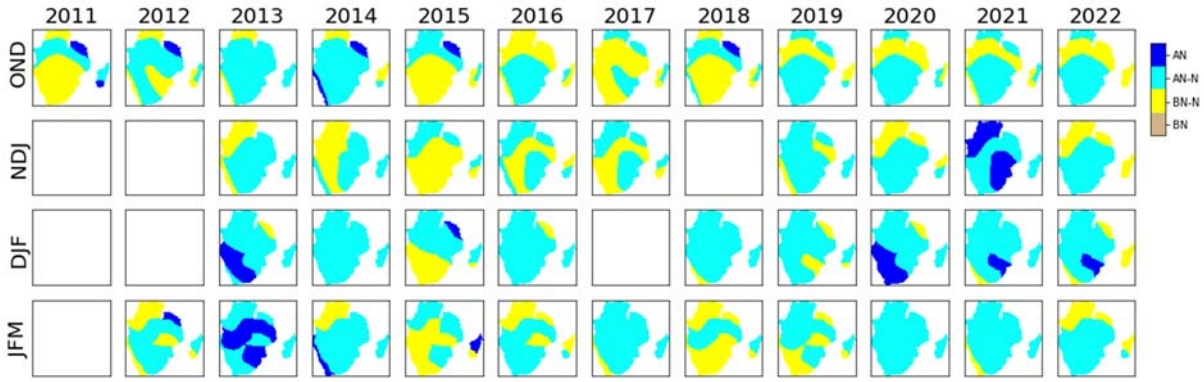


Figure 23: SARCOF forecast in the period 2013-2022. Colours denote forecast categories used in the SARCOF process and these are described in detail in this section (2.3.1.2).

SARCOF forecast remains the same since its inception in 1995, although adoption of particular tools and software 'enshrines' certain modalities of that methodology. The most recent development was the introduction in 2019 of a custom-made Climate Forecasting Tool (CFT) that replaced a series of operations with STATISTICA, MS Excel, QGIS/ArcGIS with a single software framework. The generation of a SARCOF forecast follows these steps:

1. A statistical model (CFT offers a choice of a multiple linear regression and a machine-learning model - multi-layer perceptron regression) is developed for individual stations that captures the relationship between predictands - SST or geopotential height anomalies in the month preceding the forecast, i.e. in July for the August event, and predictor - historical rainfall at

target station over the target season. That relationship is developed for a calibration or reference period, which typically is 1970-2000 (but might be 1981-2000, depending on the availability of data), for each target season (separately). To derive predictors, Extended Reconstructed Sea Surface Temperature, Version 5 (ERSST.v5) data are used, and a procedure is implemented whereby SST regions with maximum and statistically significant correlation with station rainfall are identified prior to development of the regression model.

2. That model is used to predict tercile categories for years in the validation period (i.e. the remainder of the available station observations that are not included in the model calibration). This is done separately for each target season.
3. The tercile forecast for the target season is formulated based on probabilities derived from a contingency table summarising performance of the statistical model over the validation period. The forecast tercile category is the tercile of the climatological rainfall distribution into which the line of best fit of the regression model falls into.
4. A four-category forecast for homogeneous climate zones is derived by weighted averaging of forecast categories for individual stations that fall within a zone. Four categories in the station forecast are obtained by splitting the normal category of the standard tercile forecast into two categories - below normal to normal spanning 33rd-50th percentile, and normal to above normal, spanning 50th-66th percentile. Weights reflect the level of skill that a forecast for an individual station exhibits over the validation period. Skill is measured using a hit rate (a.k.a. Probability of detection) for a dichotomous forecast. Homogeneous climate zones are defined a-prior at the level of individual countries based on similarity of interannual variability of rainfall, individually for each target season.
5. Forecast for individual stations is also derived using the IRI Climate Predictability Tool (CPT). In that MOS approach of principal component regression (PCR) is used with station rainfall as predictand and PC-transformed geopotential height of a selected dynamical forecast model as predictor.
6. The zonal forecast from CFT and station forecast from CPT, together with the evaluation of the status of modes such as ENSO and IOD (or their 'flavours') that are known to influence rainfall over analysed zones, are then synthesised to derive final zonal forecast for each of the countries. That said, there are no prescribed rules on how these various sources of information are to be combined.
7. In a discussion session with representatives of all countries, individual countries present their forecasts, and a consensus forecast is derived for regional zones, based on a relatively subjective process of amalgamation and adjusting individual country forecasts. The regional zones do not necessarily align with the country zones.

In the end, SARCOF forecast is presented as a map visualising the spatial distribution of the four forecast categories: below normal (BN), below normal to normal (BN-N), normal to above normal (N-AN) and above normal (AN). These categories have associated probabilities of the standard terciles, namely below normal (B), normal (N) and above normal (A), with boundaries between them at 33 and 66 percentile of climatological distribution of rainfall. The probabilities are fixed, i.e. they do not change from year to year irrespective of the outcome of the forecast. i.e. the BN category will always have 40/35/25 tercile (B/N/A) probabilities (Figure 24 and Figure 25). What changes between the years is only the category that is forecasted.

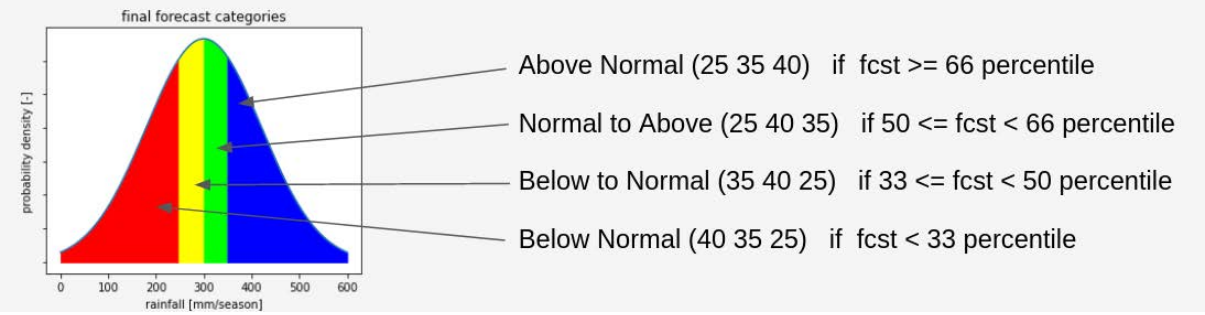


Figure 24: SARCOF forecast categories. Values in brackets represent probabilities of standard terciles. Rainfall values in the figure are for a particular location as an illustration of the principle.

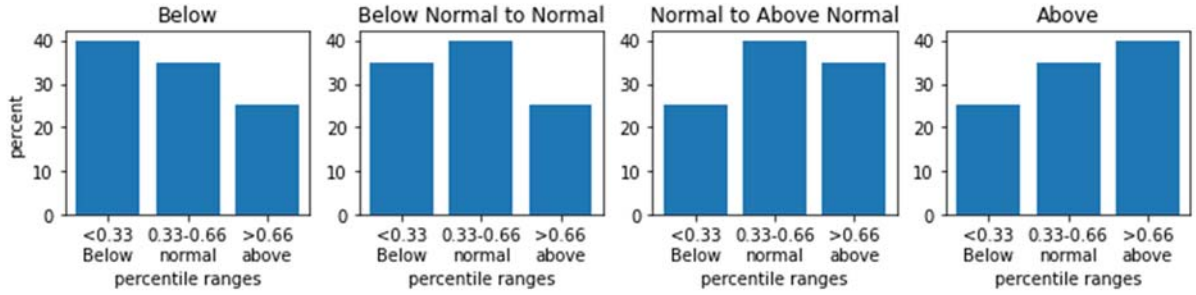


Figure 25: Fixed tercile probabilities of four SARCOF forecast categories.

2.3.1.3 Overall framework of skill analyses

As outlined in the introduction of this section, the overall framework of this activity focuses on forecasting DJF season with various lead times, considering systems from Copernicus as well as the SARCOF forecast. The skill comparison of these two systems cannot be done in an identical manner. This is because the dynamical forecasts are bound by a 6-month lead time, i.e. the earliest forecast that allows evaluation of DJF is in September. In turn, SARCOF forecast is issued only once a year in August, and has a lead time between 3 and 8 months, i.e. it covers forecasts for OND, NDJ, DJF and JFM. It is therefore impossible to compare DJF forecasts directly. Instead, the skill of the two forecast approaches is compared targeting SON and OND seasons. The skill of SARCOF forecast for DJF in absolute and not relative terms is also evaluated, i.e. without comparing it to the skill of dynamical forecasts.

In addition, because of the nature of the SARCOF forecast, whereby it is given for broad regional homogeneous climate zones (intersecting many of the considered river basins) the evaluation of SARCOF skill is presented at the grid level rather than at the basin level. The adopted grid is that of observational data used, i.e. 0.25deg grid of CHIRPS rainfall dataset.

2.3.1.4 Measures of forecast skill

WMO (Mason, 2018) recommends a range of skill measures to be used in forecast verification. Actually, an accurate probability forecast should be characterized by:

- **Reliability:** the agreement between forecast probability and mean observed frequency.
- **Sharpness:** the tendency to forecast probabilities near 0 or 1, as opposed to values near climatological occurrence probability.
- **Resolution:** the ability of the forecast to resolve the set of sample events into subsets with different outcomes.

However, for the purpose of the following analyses, only three verification metrics have been adopted (this is motivated by the need to optimize the amount of information presented in this report). These metrics are:

1. **ROC score** (ROC area under the curve): it measures the ability of the forecast to discriminate between two alternative outcomes, thus measuring resolution. It is frequently used to present the skill of categorical probabilistic forecasts, and although complex and not easy to understand, it remains one of the most comprehensive skill measures.
2. **Heidke skill score**: this score, similarly to the ROC score, reflects forecast resolution, but has the advantage of being more directly interpretable, particularly when one considers that probabilistic forecasts are often interpreted as 'deterministic' ones forecasting the most probable category. Heidke hit score measures how often did the category with the highest probability occur. Heidke skill score is a transformation of the Heidke hit score that measures the fraction of correct forecasts after eliminating those forecasts which would be correct due purely to random chance, and as such indicates an intuitive measure of forecast being better (or worse) than a random guess.
3. **Effective interest rate.** This score is one of scores that measure all or most of the important attributes of good probabilistic forecasts, rather than focusing on individual ones. WMO recommends ignorance rate over the often used RPSS, and suggests that for presentation it is transformed into effective interest rate. The effective interest rate has the advantage of being relatively intuitive and interpretable by a non-technical audience - as it provides an indication of the average returns an investor would make if they invested on the forecasts and were paid out against odds based on the climatological probabilities. This creates an intuitive measure of forecast being better (or worse) than an expectation of climatological probabilities of events.

2.3.1.5 Calibration of dynamical forecasts

For the analyses presented in this report, dynamical forecasts have been calibrated using Inflation of Variance (IoV) approach described by Johnson and Bowler (2009). This approach is akin to bias correction, and it adjusts values of individual ensemble members so that they are statistically identical to observations, where 'statistically identical' means that the ensemble members and the

observations have the same climatological variance and the same correlation with the ensemble mean. The approach ensures reliability of the forecast, i.e. that the spread of observations is adequately 'sampled' by the forecast (under the assumption that the spread of observations can be described by normal distribution, which is fulfilled at seasonal time scale). The method is frequently used to calibrate operational seasonal climate forecasts (e.g. Cottrill et al. 2013) and hydrological forecasts (Roulin E., & Vannitsem S, 2015), and relatively simple to implement.

More specifically, the calibration of the forecast has been done with respect to CHIRPS blended observations-satellite rainfall data product (Funk et al. 2015). That dataset is frequently used in a range of applications over southern Africa. This method has been implemented on a monthly basis using a leave-one-out approach, i.e. data for each month/lead time and each year are calibrated independently using parameters derived from all data for that given month/lead time with that year excluded. This approach effectively ensures that when the calibrated data are used in the process of model verification to calculate forecast skill, the outcome is a cross-validated skill measure.

2.3.2 Results

2.3.2.1 Skill of SARCOF forecasts

SARCOF forecast shows a relatively high skill measured by ROC score for below normal and above normal terciles (reaching values of 0.85-0.9), but low skill for the normal tercile (Figure 26). The high skill for the two 'non-normal' terciles is not accompanied by high skill in the other skill measures, i.e. in the effective interest rate (EIR) and Heidke skill score (HSS). These two skill measures display relatively low levels of skill, scattered in spatial 'pockets'. There appears to be some (although relatively weak) reduction of skill with lead time in ROC for 'non-normal' terciles, but an increase in skill with lead time for the normal tercile. EIR and HSS seem not to show any change in skill with lead time.

While the overall impression on skill of SARCOF forecast is optimistic, Figure 26 indicates that the forecast has a very poor reliability and, in particular, the below normal tercile has never been forecasted as a tercile with the highest probability. The explanation of the apparent contradiction between that lack of forecast of the below normal tercile (Figure 26) and the obviously high ROC skill for that tercile (Figure 26). lies in the fact that the forecast of the SARCOF 'N-AN' category gives a probability of below normal tercile at 35%, which is higher than the climatological probability and, hence, that category is forecasted approximately 50% of the time. The same applies to the above normal tercile. By extension, the lack of skill in the normal tercile results from the fact that this tercile has always a probability that is higher than climatology, independent of which SARCOF category is forecasted. As a result, the forecast of the normal tercile is bound to fail two out of three times (as the normal tercile occurs only once in three years on average, resulting in no skill of such forecast.

The two remaining measures of skill, i.e. EIR and HSS are free of the artificial inflation of skill resulting from the overlay of tercile probabilities and the SARCOF category forecast frequency. As mentioned earlier, these measures express low levels of skill, with little spatial coherence. The spatial structure of these two skill measures does not seem to correspond to known spatial patterns of influence of

global modes of climate variability, two of which are illustrated in Figure 28, and neither does it bear any resemblance to the distribution of homogeneous climate zones used to present SARCOF forecast.

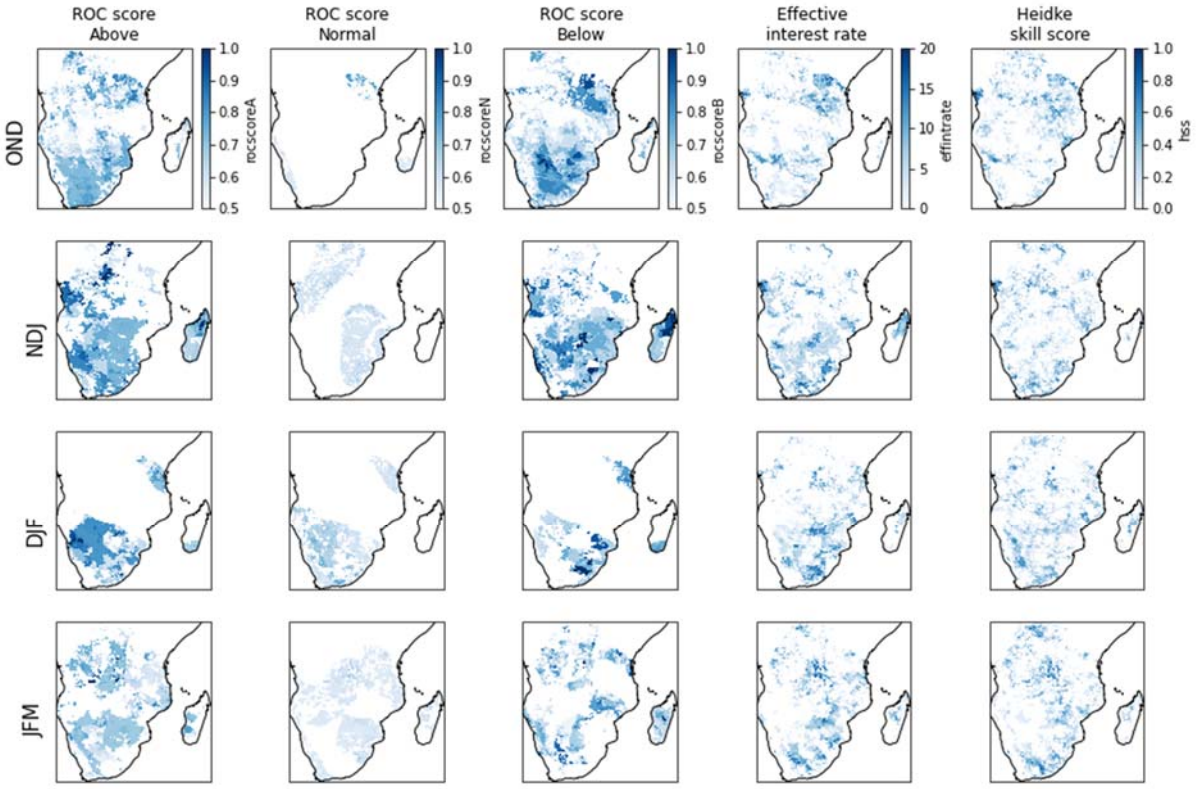


Figure 26: *S*kill of SARCOF forecasts.

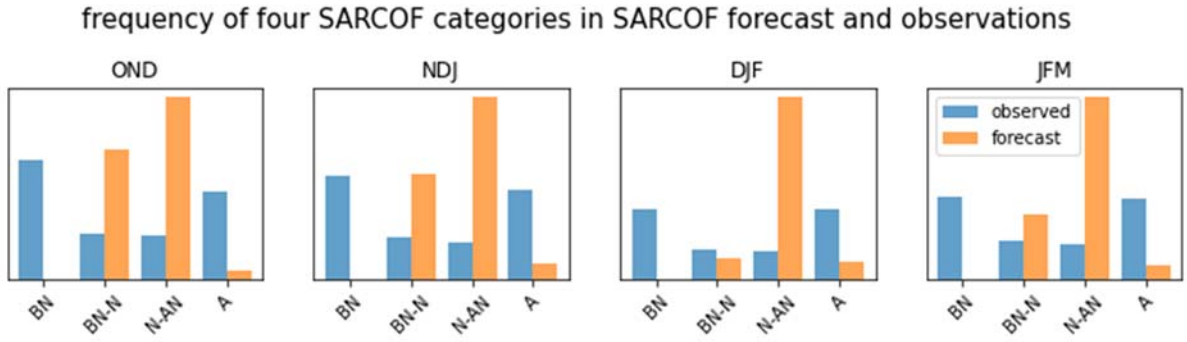


Figure 27: Frequency of the four SARCOF forecast categories in SARCOF forecast and observations over overlapping periods over all locations in the SADC region. Note that frequencies of the two-central categories in observations are lower than those of the outer categories because the latter are terciles, while the former split the central tercile. A reliable forecast would have frequencies of forecast categories corresponding to frequencies of observations.

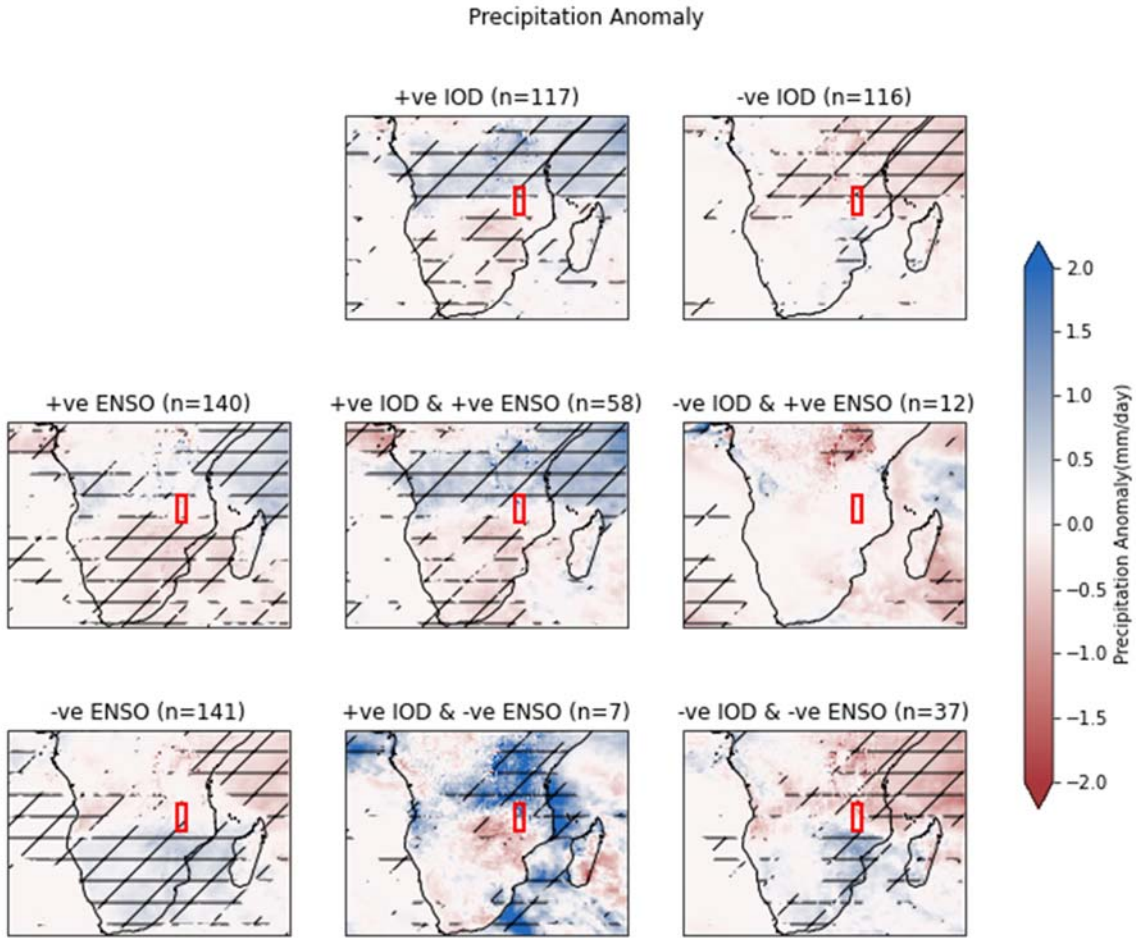


Figure 28: Patterns of rainfall anomalies (ERA5 rainfall) associated with anomalies in ENSO and IOD - the two main modes of climate variability known to influence southern African rainfall and potential sources of seasonal predictability.

2.3.2.2 Skill comparison of SARCOF and dynamical forecasts

As outlined earlier, SARCOF and dynamical forecasts skill can be compared only for forecasts issued in August for OND and NDJ seasons. Relative frequency - a measure of the forecast reliability (i.e. ability to forecast events with observed frequency) indicates that while SARCOF forecast is clearly biased towards the normal category, the dynamical systems generate forecasts that are relatively uniformly distributed across all three terciles for both OND and NDJ forecasts (Figure 29).

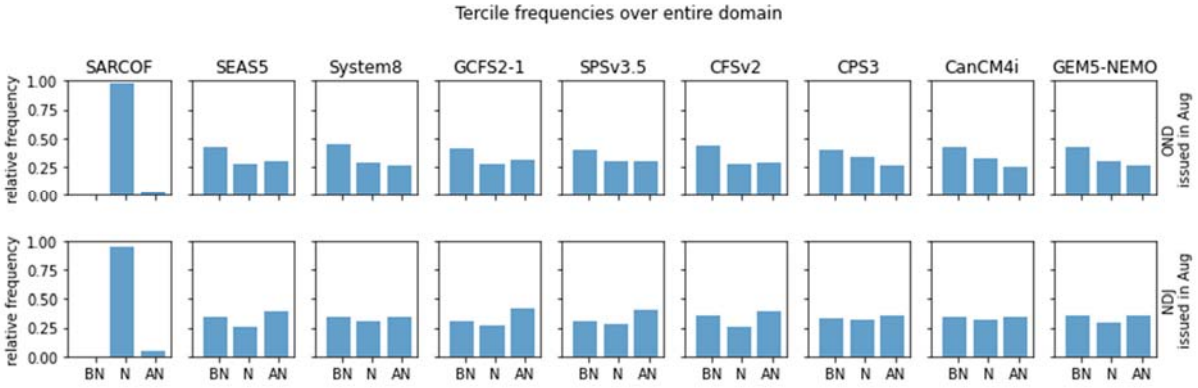


Figure 29: Relative frequency of forecasted tercile (tercile with maximum probability) in SARCOF and dynamical forecasts. Frequency represents all grid points in the domain overall years for which the forecast is available.

In terms of forecast skill, SARCOF forecast shows relatively high ROC score for below normal and above normal terciles over Tanzania and summer rainfall region of South Africa in OND, and South Africa, Botswana, Zimbabwe and Angola in NDJ. However, as argued above, this skill seems to be a spurious byproduct of the combination of tercile probabilities used by SARCOF forecasts and bias of the forecast towards the normal tercile and does not reflect a real skill.

The dynamical forecasts show considerably lower levels of skill, with only SEAS5 and GEM5-NEMO systems depicting higher values over Tanzania in OND (that manifests in each of the skill measures used here, apart from ROC score for normal tercile as seen in Figure 30). CFSv2 and CPS3 systems show a similar pattern of skill to the former, but with lower values. The remaining models show weak skill with scattered spatial patterns.

On the other hand, forecasts for the NDJ season have substantially higher levels of skill than forecasts for OND (Figure 31). The four systems mentioned above show skill over Tanzania, and three of those (SEAS5, CFSv2 and CPS3) show high skill over southern Africa south of -15 deg latitude. Skill in that part of southern Africa is also displayed by SPS3.5 and GCFSv2.1. Those skill patterns are consistent across the various skill measures apart from the ROC score for normal tercile. Similarly to OND forecast, this skill metric shows lower values and scattered spatial patterns.

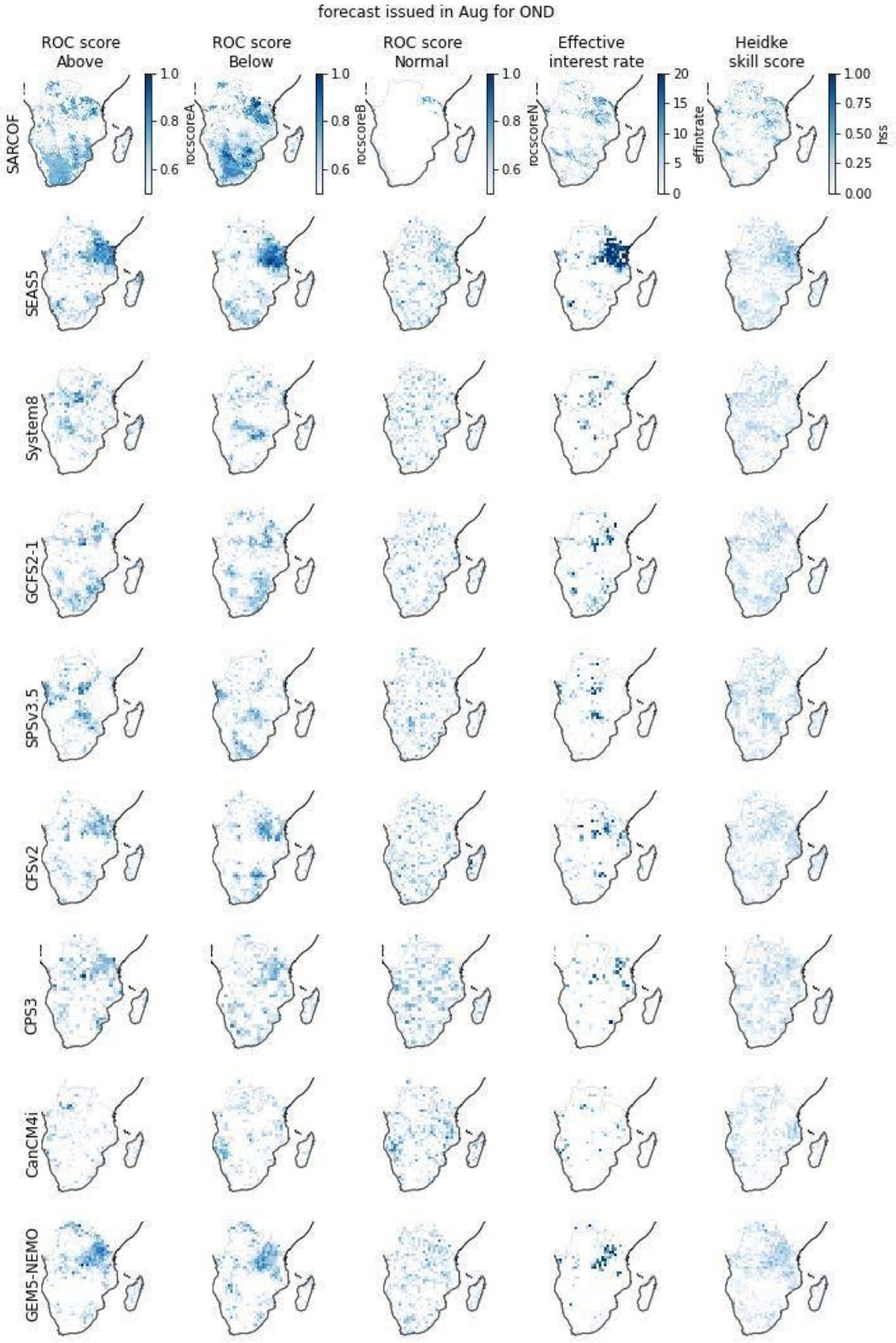


Figure 30: Skill of SARCOF and dynamical model forecasts of OND rainfall issued in August.

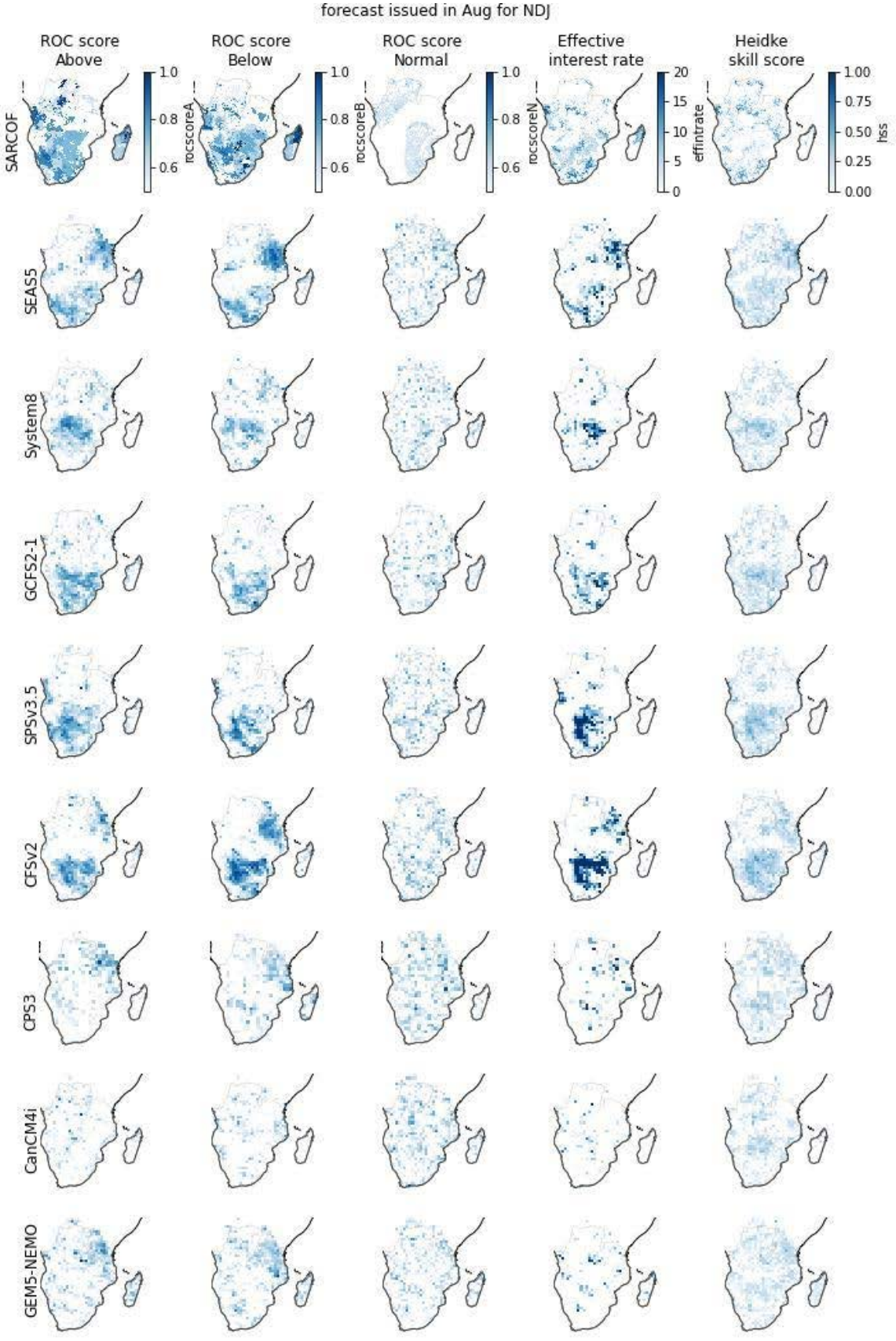


Figure 31: Skill of SARCOF and dynamical model forecasts of NDJ rainfall issued in August.

2.3.2.3 Skill of dynamical models in forecasting DJF rainfall in the hydrological basins

The skill of the DJF rainfall forecast for river basins issued in Sep-Dec is visualized in the maps of Figure 32,Figure **33**Figure 34Figure 35 and bar charts for the below normal tercile ROC score (Figure 36, Figure **37**, Figure **38** and Figure **39**). In bar charts, the individual basins are clustered within the 15 main river basins of the region illustrated in **Figure** 22. These figure show that there is only limited skill consistency across basins, forecasting systems and lead times, and while the figures provide means for selecting skillful forecasting systems for individual basins (save the consideration of skill arising by chance) it is difficult to derive broad generalizations from them. However, some key features can be highlighted:

- In general, main river basins for which forecasts are relatively skillful across lead times and forecasting systems include: Tanzanian East Coast, Limpopo, Okavango, parts of Orange and parts of Zambezi.
- There is in general low skill and low consistency in basins in coastal Angola, Congo, Madagascar, and south-east coast of South Africa.
- Considering context provided by the Malawi case studies, DJF rainfall forecast for Shire shows skill only for forecasts issued in Dec.

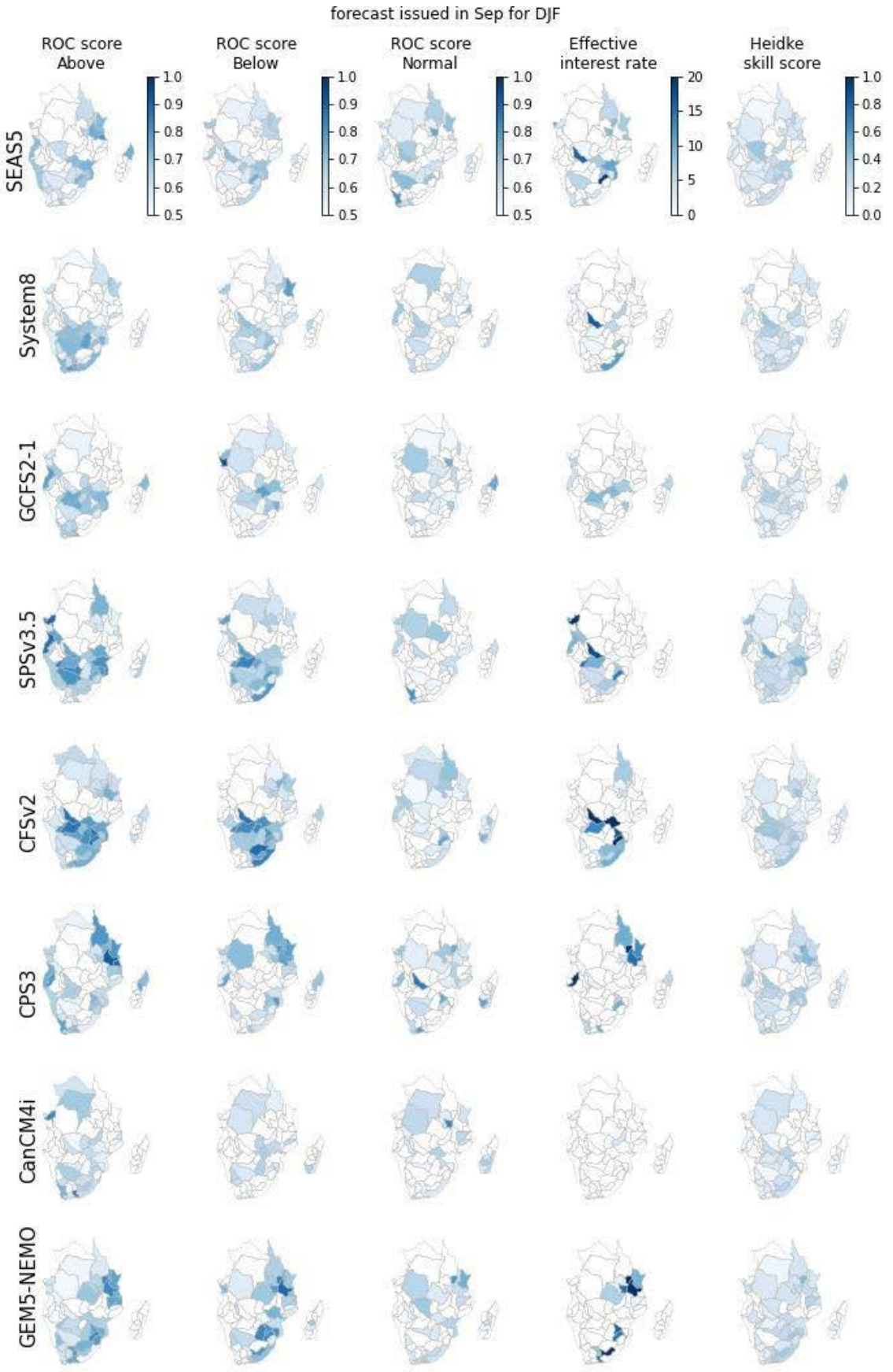


Figure 32: Skill of forecast of DJF rainfall in river basins issued in September.

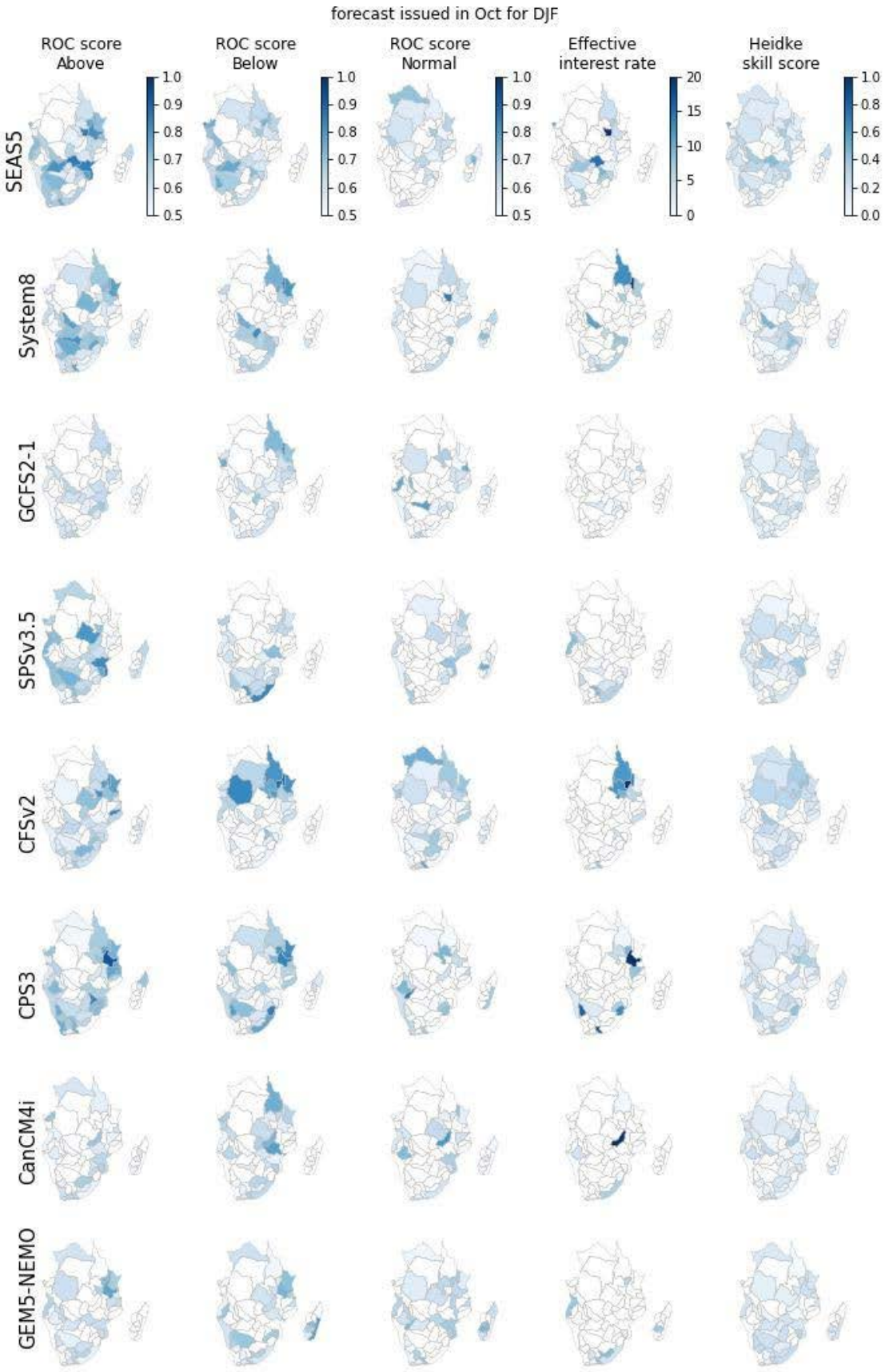


Figure 33: Skill of forecast of DJF rainfall in river basins issued in October.

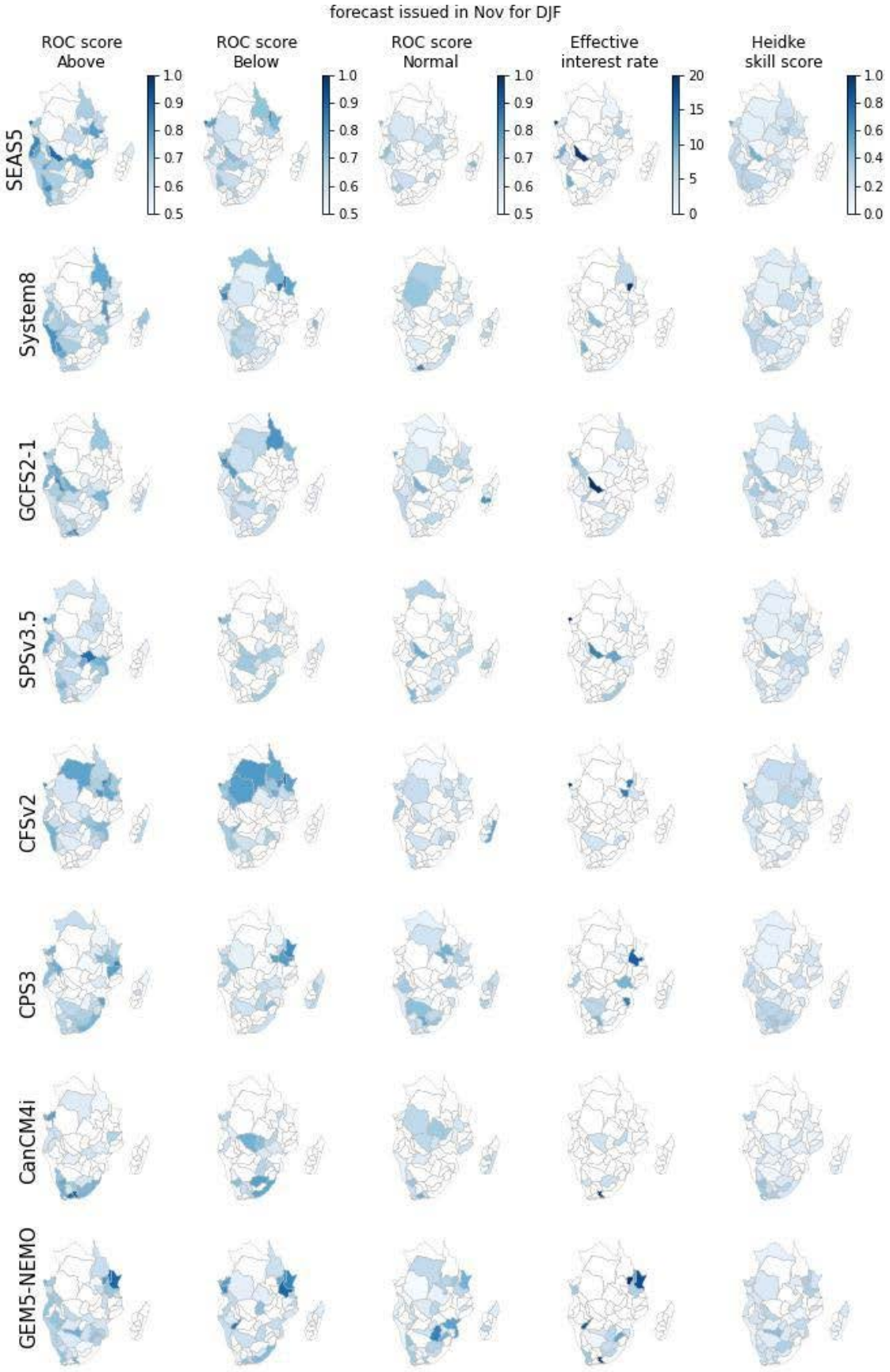


Figure 34: Skill of forecast of DJF rainfall in river basins issued in November.

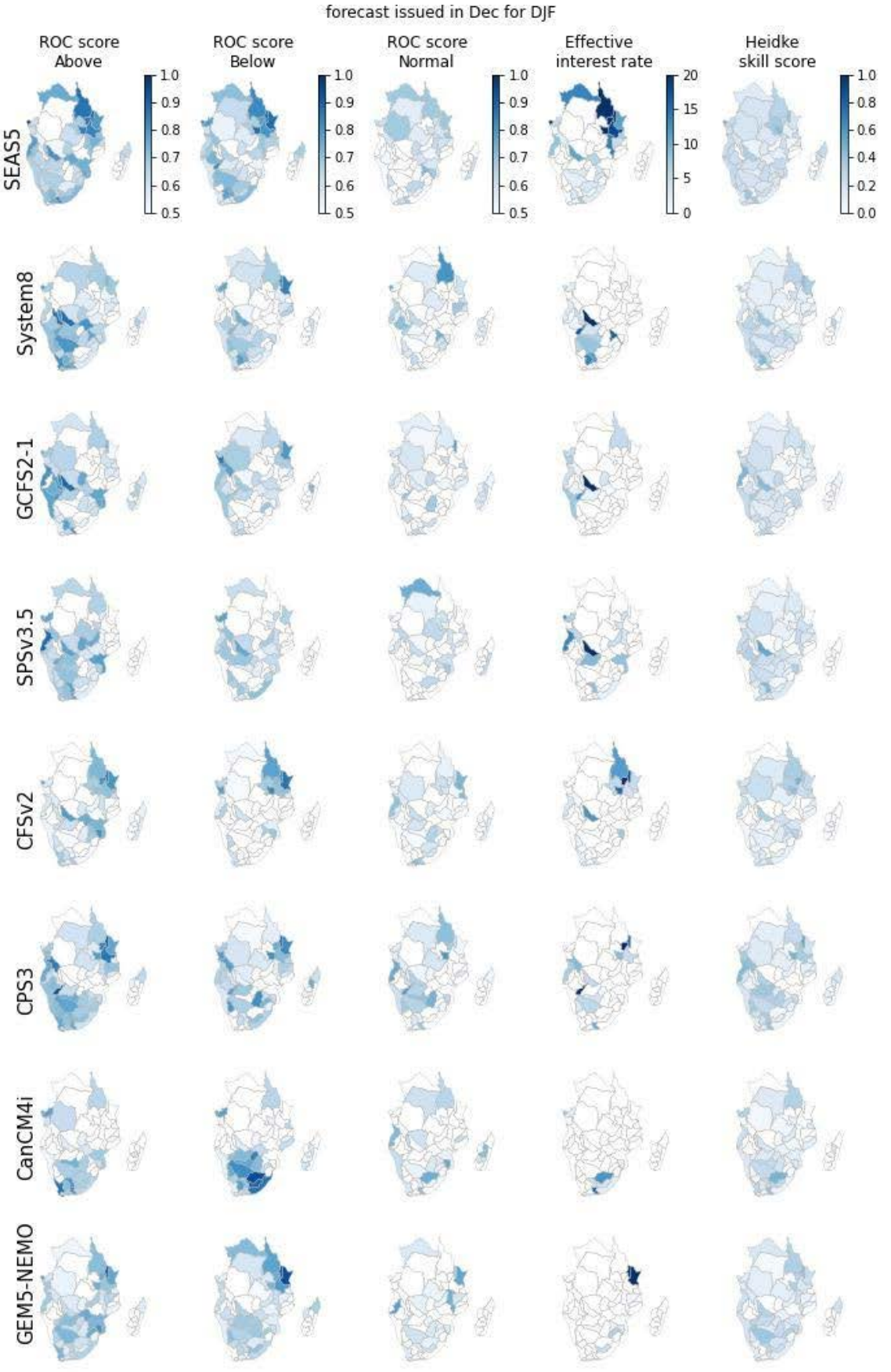


Figure 35: Skill of forecast of DJF rainfall in river basins issued in December.

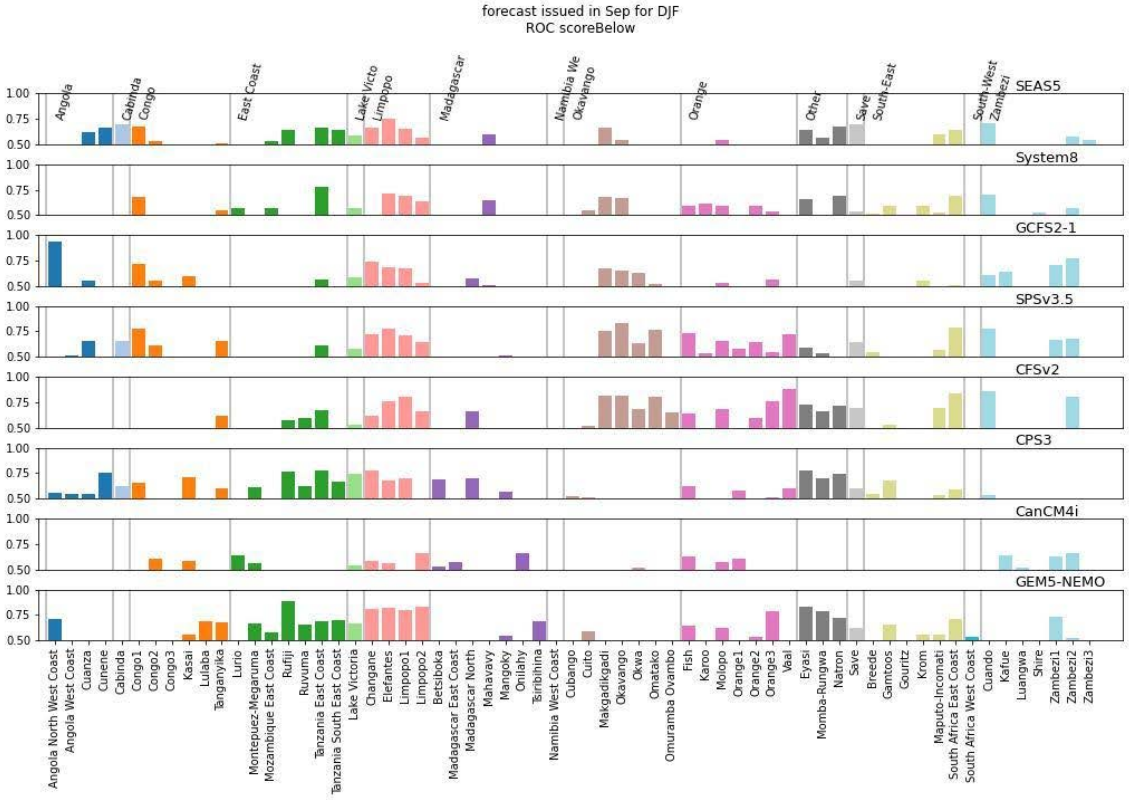


Figure 36: Skill of forecast (ROC score for below normal tercile) of DJF rainfall in 63 SADC river basins issued in September.

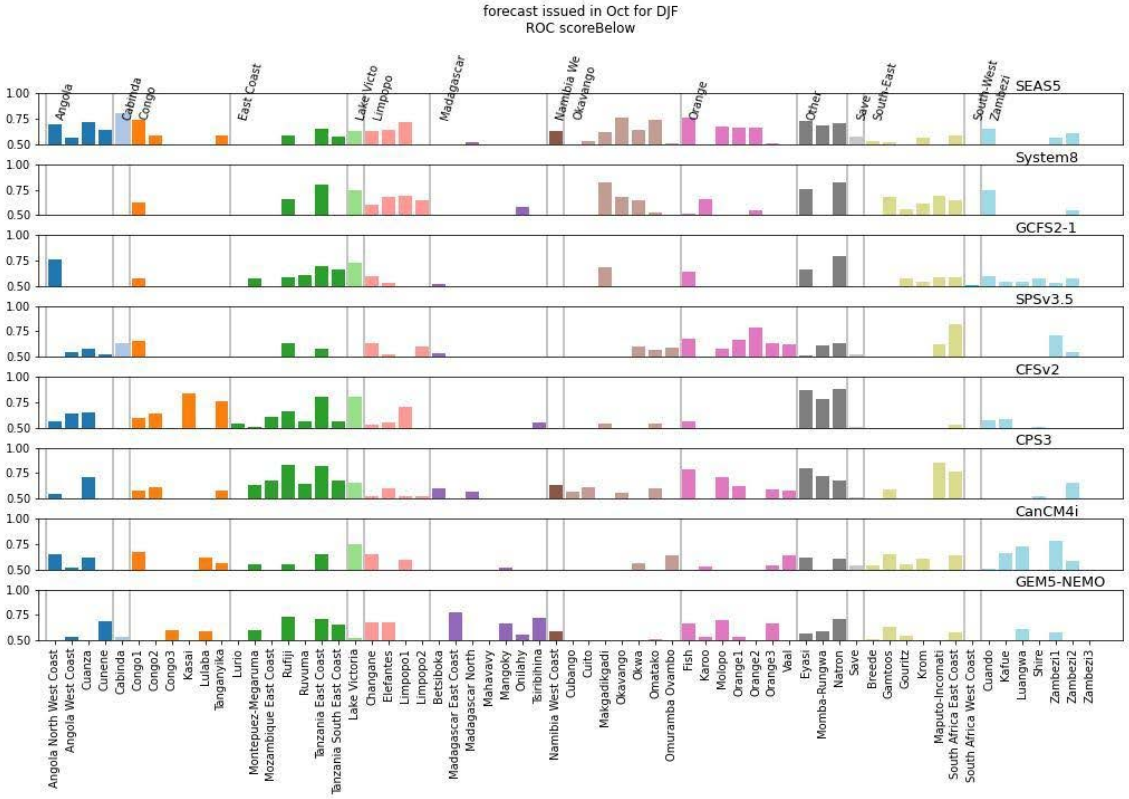


Figure 37: Skill of forecast (ROC score for below normal tercile) of DJF rainfall in 63 SADC river basins issued in October.

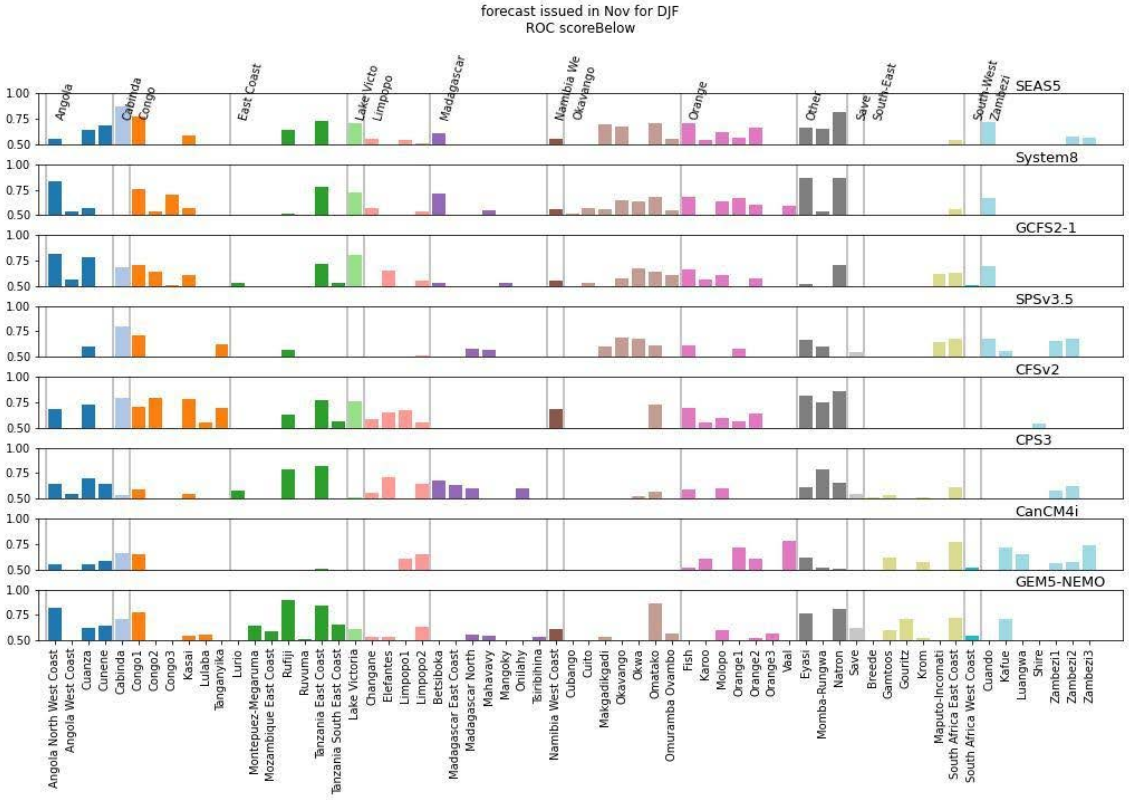


Figure 38: Skill of forecast (ROC score for below normal tercile) of DJF rainfall in 63 SADC river basins issued in November.

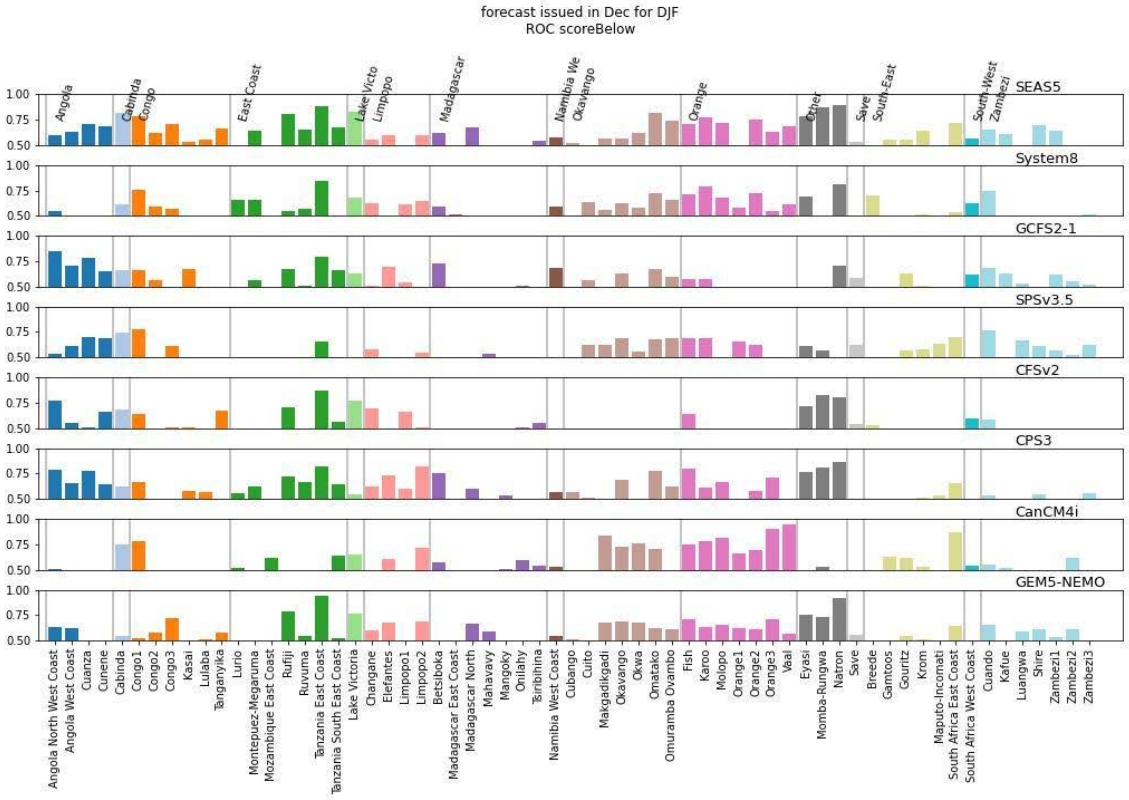


Figure 39: Skill of forecast (ROC score for below normal tercile) of DJF rainfall in 63 SADC river basins issued in December.

2.4 Review of seasonal forecasting approaches at SARCOF and SWIOCOF

The review of the climate outlook forums (SARCOF and SWIOCOF) has been conducted in parallel to the characterization of climate services' gaps in WP7. The process has followed a competency-based analysis considering five competencies: i) create and manage climate data sets; ii) derive products from climate data sets; iii) create and interpret climate forecasts and models; iv) ensure quality of climate forecasts; and v) communicate information to users (as described in the WMO Guide to Competency (WMO, 2018).

All SADC National Meteorological and Hydrologic Services (NMHSs) participate in the SARCOF process and communicate the seasonal forecast during their National Climate Outlook Forum (NCOF). The IRI Data Library, Climate Prediction Center (CPC), National Oceanic and Atmospheric Administration (NOAA), and United Kingdom Meteorological Office (MO), are the principal climate predictor's datasets used by these services. They are provided at monthly time scales and in Network Common Data Form (NetCDF) format.

During previous SARCOFs, SYSTAT7 was used for statistical forecasting, but the method was improved with the development of the Climate Forecasting Tool (CFT) by SADC-CSC during the Satellite and Weather Information for Disaster Resilience in Africa (SAWIDRA) project. This tool was released during SARCOF 23 in August 2019. Initially, it used Linear Regression (LR), but subsequently in the following SARCOF 24 it was updated with Artificial Neural Networks (ANN), which uses Multilayer Perceptron (MLP) regression.

During previous SARCOFs, only the SST predictor was used together with related climate drivers for that year, however, for the past two years more variables like MSLP, temperature at 2m, wind speed and direction at different levels, among others, were introduced and are currently being used during SARCOF and NCOF's. In SWIOCOF, the seasonal forecast uses the number of cyclones, precipitation, the accumulated cyclone energy and number of cyclone days during the season as predictands. Dynamical approach is also considered based on global operational forecasts products available at ACMAD and Regional Specialized Meteorological Centre (RSMC) La Reunion. The SWIOCOF put emphasis on statistical-dynamical approach because of its improved forecasting skills. The second tool used at SARCOF is the CPT, which has been in use since the beginning of the SARCOF process. During SARCOF 24, more training was done on the tool, focusing on enabling experts to conduct station forecasts, and use more predictors from the CPC data set. Seasonal North American Multimodel Ensemble (NMME) hindcast and forecast were also introduced and used for the first time.

Regarding verification, SADC-CSC trained a few experts on the use of a verification template, which includes probability of detection, bias skill scores, hits and false alarms. The methodology uses observed data, which is transformed in the tercile scale, using the long term mean and the standard deviation based on the long-term average. There was also an evaluation of the SARCOF rainfall outlooks from 2001 to 2012 done by FEWSNET (Magadzire, 2012) to determine the performance of the forecast with the observations. The evaluation showed that the SARCOF forecasts over those twelve years on average performed very well. SADC-CSC had an improved version of the verification of the SARCOF 22 rainfall outlook, which used Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) version 2 and included the half hit and half miss approach. During SARCOF 24,

in August 2020, ACMAD introduced and trained SADC-CSC officers on the ranked probability skill score (RPSS) method, but the knowledge was not cascaded down to the NMHS climate experts in the region. At that same SARCOF, SADC-CSC used an improved method of verification, whereby the climate drivers were considered. From the information gathered it can be concluded that there is limited exposure of the experts to verification methods and tools. The methodology for intercomparison and comparison of operational global or regional forecasts with observations and the user-relevant verification schemes are hereby identified as regional capacity gaps to be addressed using WMO guidelines and manuals for available best practices.

In these SARCOF processes, before the seasonal forecast is released, there is a presentation on the 'Current Status of the Global Climate System' which is usually presented by a climate expert seconded to SADC-CSC. Mostly the key drivers examined are the current SST's, the current and forecasted El Nino Southern Oscillation (ENSO) conditions, the Indian Ocean Dipole (IOD), the Southern Indian Ocean Dipole (SIOD), and conditions over the South-West Indian Ocean (SWIO). These are then combined with the global model rainfall outlooks from sources such as IRI, National Centers for Environmental Prediction (NCEP) and the South African Weather Service (SAWS). Professionals from other partners like the Famine Early warning Systems Network (FEWSNET) and from different climate sensitive sectors also make presentations from research that they have done, or on the impacts of the previous rainfall season. Usually after SARCOF, experts present their results to their relevant NMHS, including the skills of the different models and predictors used.

In fact, in the SADC region, monthly agro-meteorological and drought monitoring bulletins are produced by NMHSs, but they do not reach all relevant users. SADC-CSC provides information on the institutional website, while some SADC countries also have their own websites. However, most of the SADC NMHSs are still in the process of developing an effective and user-friendly online platform. Seasonal Outlook Statements are communicated using emails, webpages, and various media platforms (radio, newspaper, TV, social media etc.). SADC-CSC also prepares an early warning bulletin for climate sensitive sectors, which is loaded on the SADC-CSC website, an example of which can be seen for the 2018/19 rainfall season.

Apart from the sector breakout groups during the SARCOF and the production of the Early Warning Advisory Bulletin, little is done in relation to tailoring climate information to specific user needs by co-designing and co-developing products with users. From the survey component on climate services communication, most SADC members (more than 50%) are lagging in terms of GFCS full implementation. However, the main gap identified under seasonal forecasting is the limited use of GCM data sources in the region. There is not good knowledge on how to retrieve data from the different data sources and there is little training done on applications like CPT, and on the documentation, communication, and interpretation of the skills of the model outputs. Intra seasonal drivers of seasonal climate (e.g. tropical/equatorial waves) are not considered in the seasonal forecasting approach in SARCOF/SWIOCOF resulting in poor forecast in years when intra seasonal drivers dominate the seasonal climate variability. Besides, the evaluation of seasonal forecast quality through the forecast quality assessment should also be considered. This involves setting up a quality assessment framework to provide end-users with the tools to understand which approaches could better fit their interests. Showing the skill score of a forecast to the user increases the reliability and

usefulness of the forecast. Such skill scores will vary depending on the season, the variable being evaluated, the region as well as the lead time.

2.5 Review of seasonal forecasting approaches for the rainfall season in Tanzania

In Tanzania one of the most important climatic features is the alternation of dry and wet seasons (due to its impacts on food security and farming). The dry season comprises May to September months, whereas the wet season has two presentations:

- **Bimodal:** in this mode there are two rainy seasons, March to May (Masika rains) and October to December (Vuli rains). This behaviour is found over the lake Victoria Basin, the northeastern highlands and northern coasts of the country.
- **Unimodal:** in this mode there is only one rainy season, from November to April (Msimu rains). This behaviour is found in southern, western and southwestern Tanzania.

In this framework, the National Weather and Hydrologic Service of Tanzania (Tanzania Meteorological Agency, TMA), provides different types of forecast information related to rainfall in the country, both for bimodal and unimodal areas. These predictions are issued considering the specific season of interest for each region. For instance, the Oct-Dec forecast was issued in late August, while in September, TMA issued the forecast for Nov-Apr season. Figure 43 Figure 49 review this information.

Firstly, the most recent predicted season is verified against observations, both regarding the precipitation amount, (Figure 40) and the onset of the rainy season (Figure 41). Since the forecast is issued in zones, the overall performance is averaged over the country. With regard to the term accuracy used in the report (Figure 40), it is computed as a matching between forecasts and observations (expressed in percentage). More specifically, TMA uses Probability of Detection (POD) or Hit rates, which is the ratio of the number of correct forecasts to the total number of correct forecasts and missing events (from a contingency table).

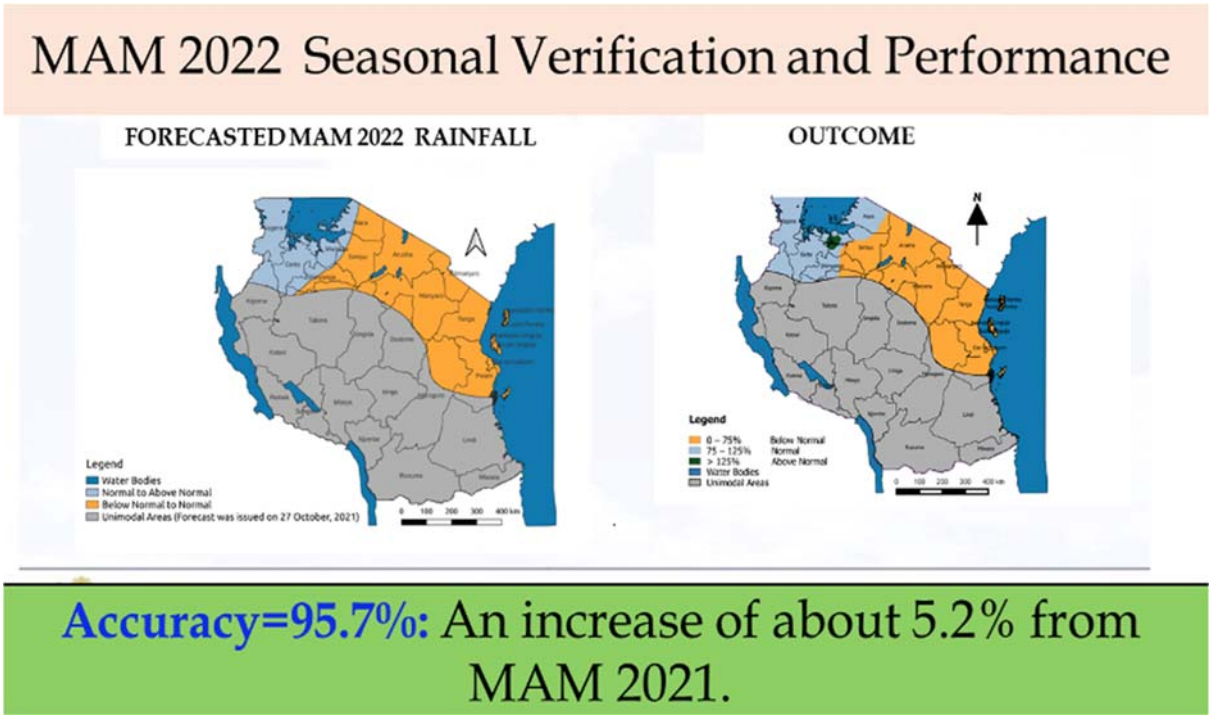


Figure 40: Comparison of the seasonal forecast for MAM 2022 and the observed outcome. In green, performance comparison with the former MAM 2021 prediction. The grey colour corresponds to the unimodal rainfall regime area, whereas the rest is for the bimodal regions.

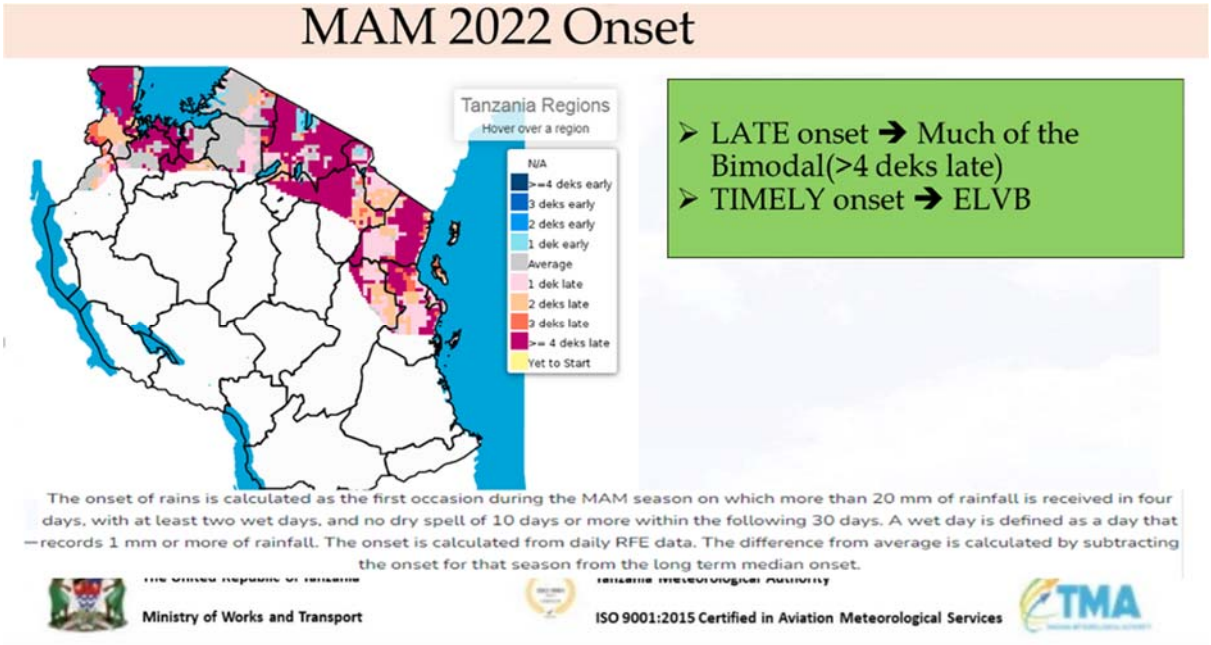


Figure 41: Verification of the onset of the rainy season for MAM 2022 in the bimodal areas. Each 'dek' corresponds to 10 days. ELVB is the acronym for East Lake Victoria Basin.

After this revision, the prospects for the next season are set by inspecting the status and prediction of the El Niño Southern Oscillation and the Indian Ocean Dipole. Figure 42 to Figure **45** show this information for the OND 2022.

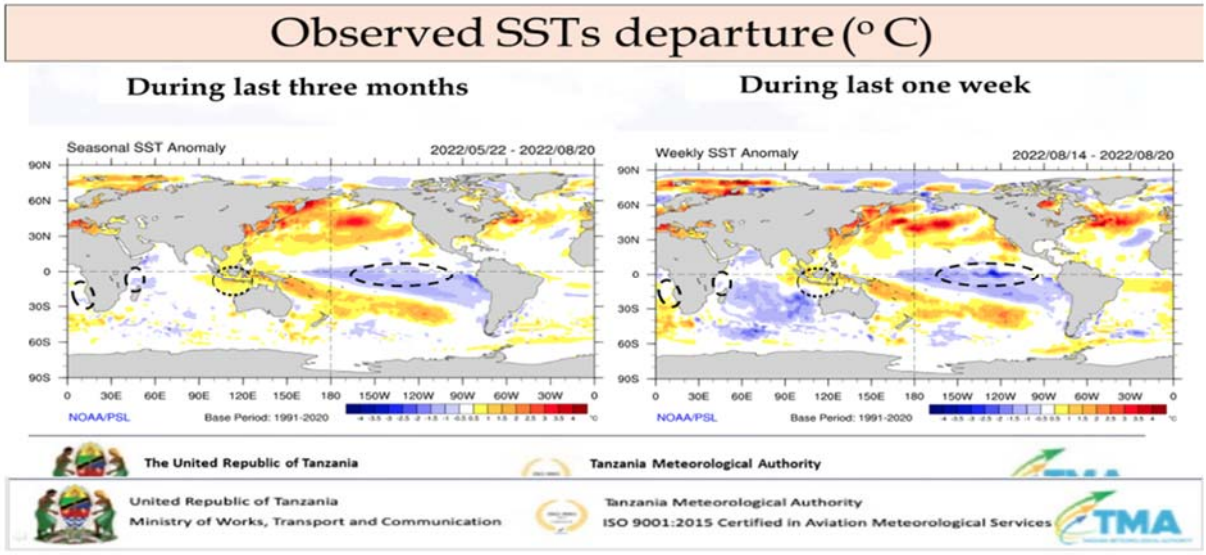


Figure 42: State of the sea surface temperature (SST) Anomalies in the different relevant regions for the Tanzanian climate. The left figure is for the last three months, whereas the right one corresponds to the anomalies in the most recent week.

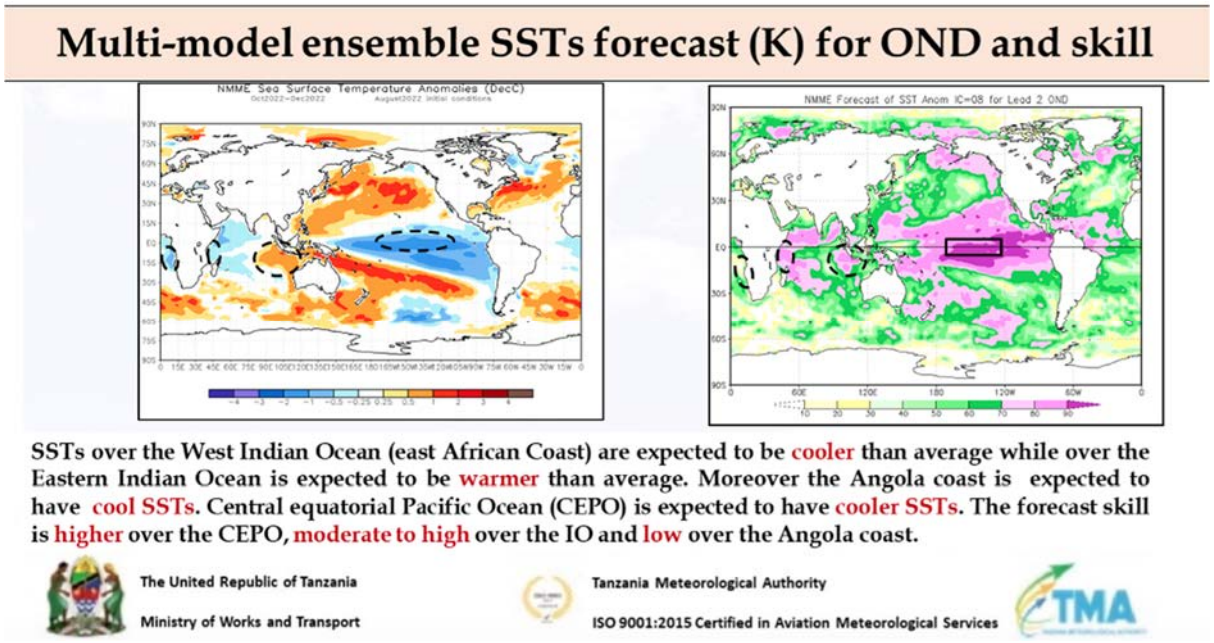


Figure 43: Sea surface temperature seasonal forecast for the OND season (left), with the associated skill (right).

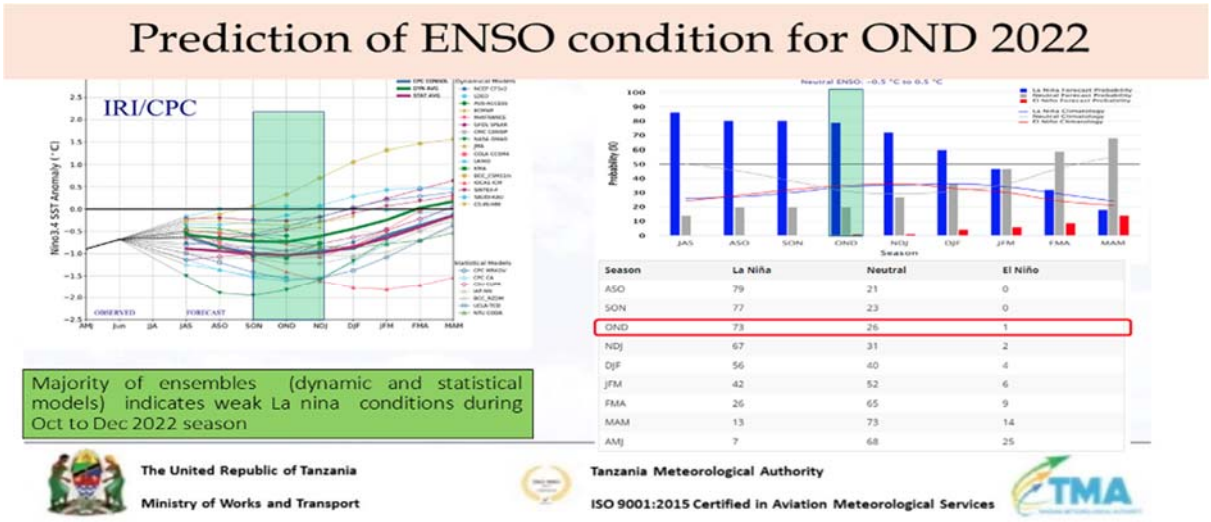


Figure 44: Prediction for OND 2022 ENSO conditions.

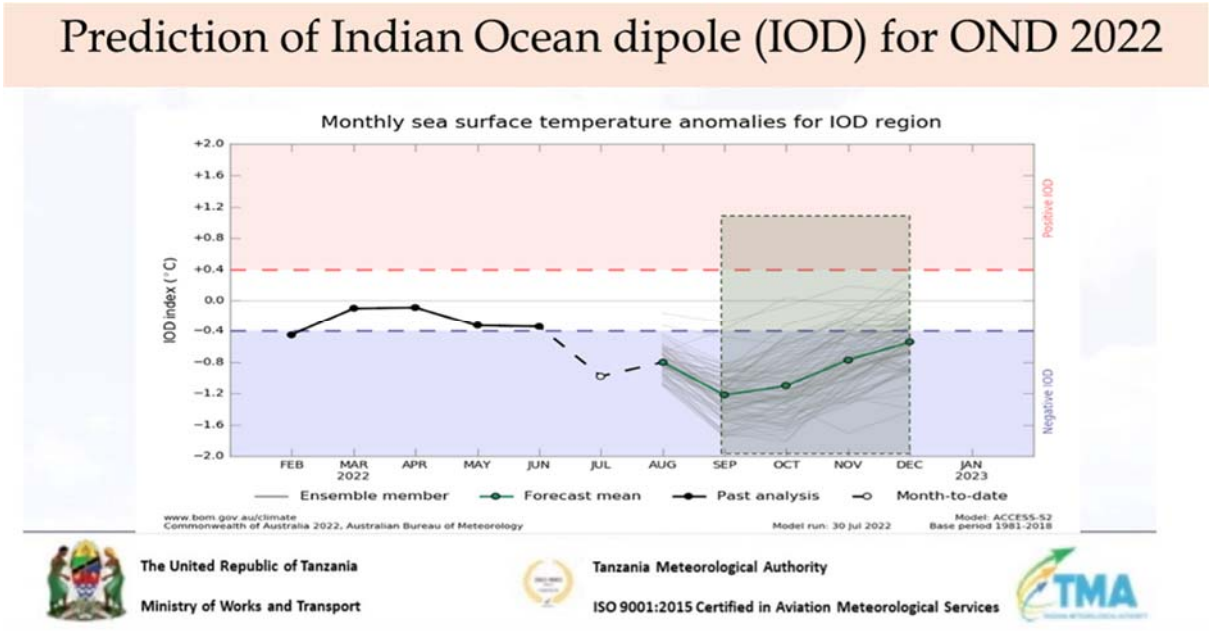


Figure 45: Seasonal prediction of the Indian Ocean Dipole for the following months.

Finally, the seasonal rainfall outlook for the upcoming season is issued (amount and onset). An example for OND 2022 is provided in Figure 46 Figure 49.

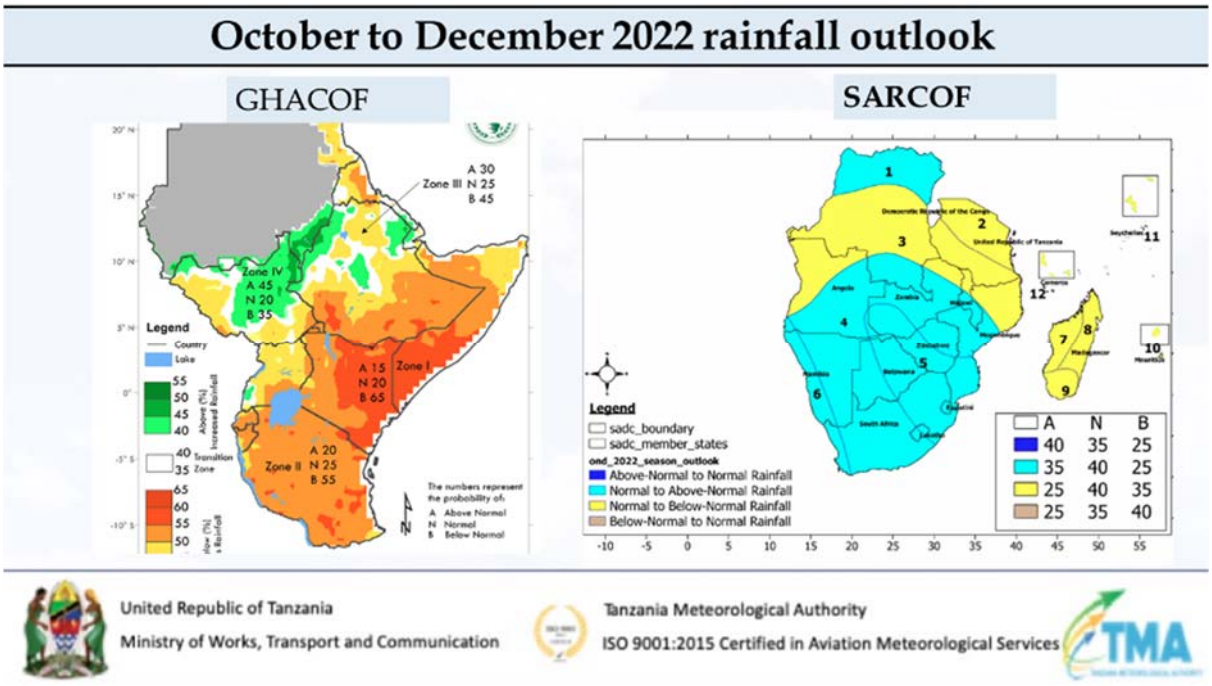


Figure 46: October to December rainfall outlook from the GHACOF (left) and SARCOF (right).

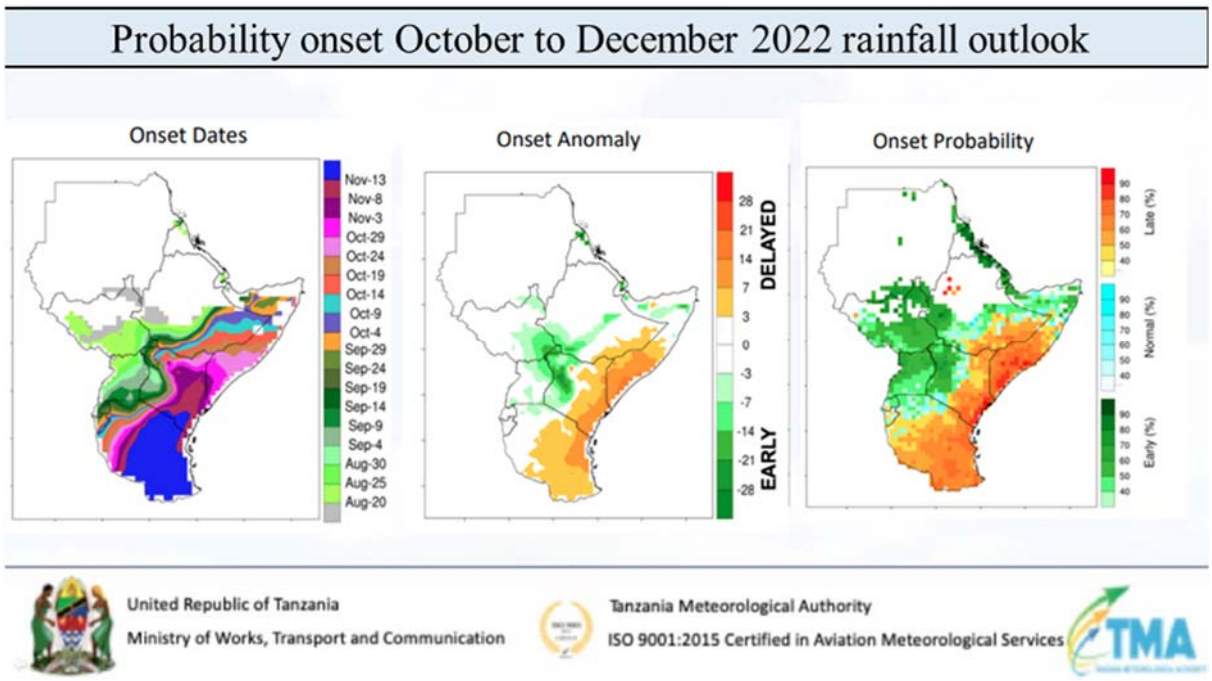


Figure 47: Seasonal forecast of the rainfall season onset (left), with the associated anomalies (centre) and probability (right).

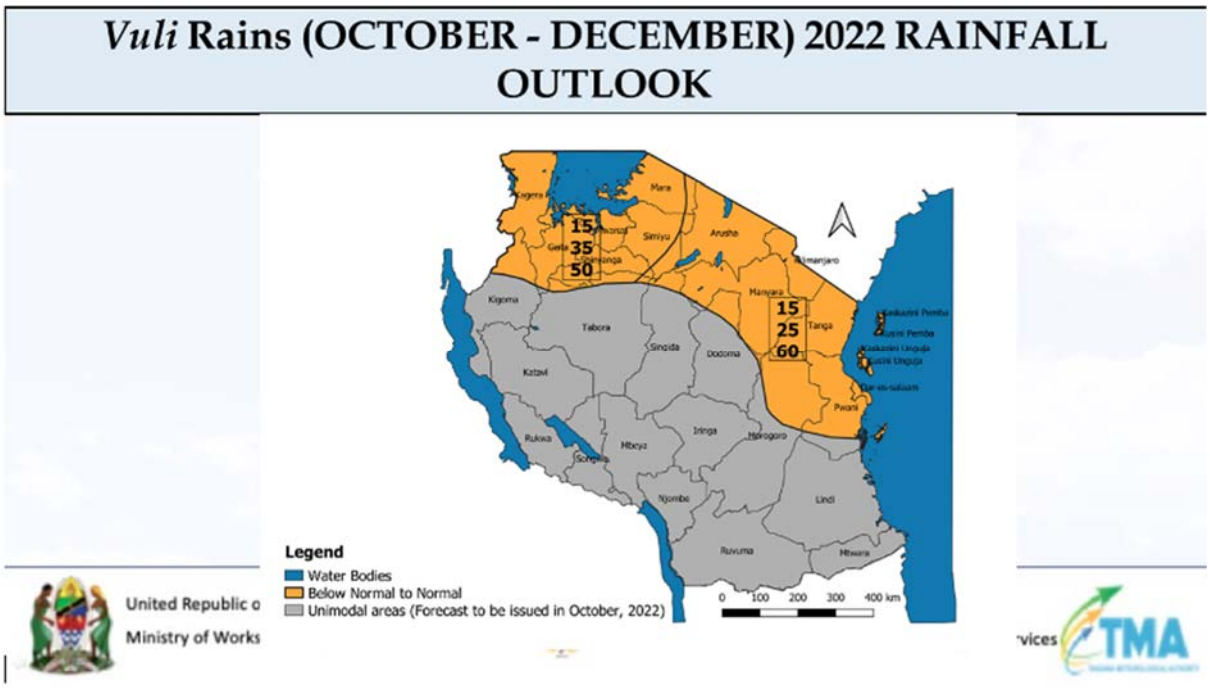


Figure 48: Rainfall outlook for OND 2022. The grey colour corresponds to the unimodal rainfall area, whereas the rest is for the bimodal regions.

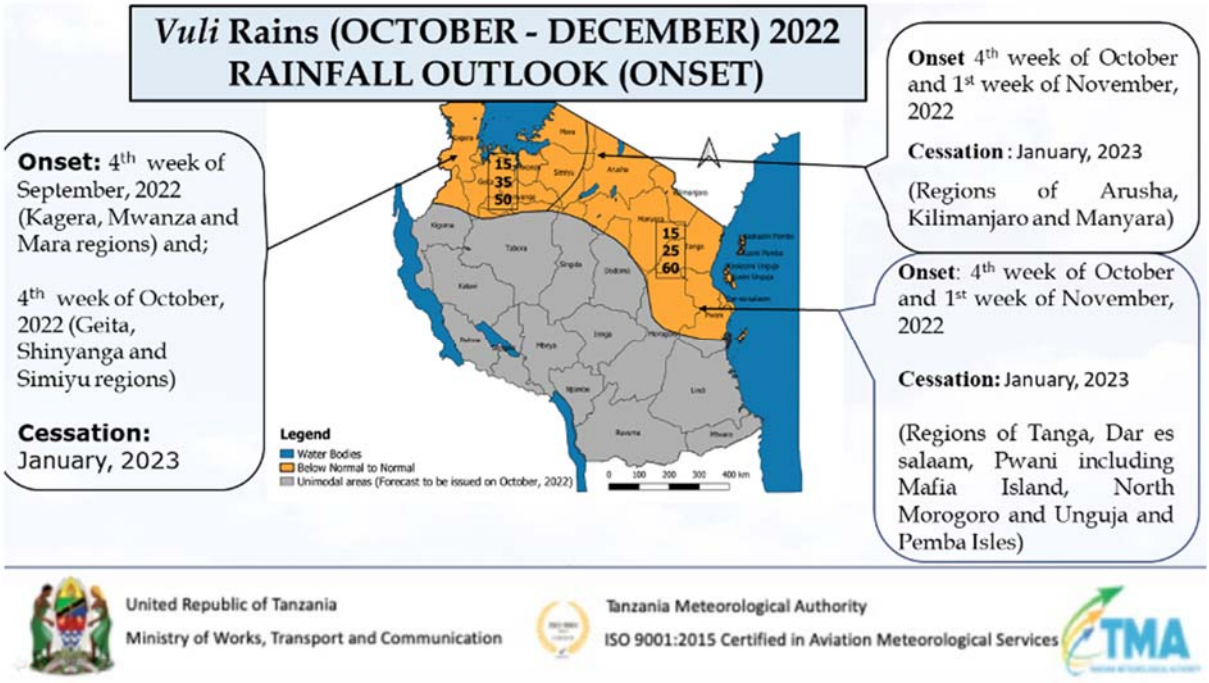


Figure 49: Rainfall outlook for OND 2022 rainy season with comments on the onset and cessation. The grey colour corresponds to the unimodal rainfall area, whereas the rest is for the bimodal regions.

3 Decadal prediction

Decadal climate predictions are a source of climate information for inter-annual to decadal time scales, filling the gap between seasonal climate predictions and climate projections. There is an increasing interest of users in having information on how the climate system will evolve during the following years. Typically, forecast products are issued for annual, multi-annual, and multi-seasonal averages, depending on specific user needs.

The predictability on these time scales is given by both the external forcings (both natural and anthropogenic) and the internal climate variability (natural slow variations of the climate system, Doblas-Reyes et al., 2013; Goddard et al., 2013; Smith et al., 2019). Decadal predictions are produced with the same forecast systems as the climate projections. The only difference between them is that, in addition to the information about external forcings (which is used to produce the climate projections), the decadal predictions are run from a set of initial conditions created with the observed climate state. With this procedure, known as model initialization, the decadal predictions aim at being in the same phase as the actual climate variability (Hazeleger et al., 2013). After some forecast years, the forecast systems lose the information about the initial conditions, and the evolution of the decadal predictions converges to that of the climate projections.

3.1 Data and methods

The forecast quality assessment has been applied to spatial fields of near-surface air temperature, sea level pressure and precipitation forecasts from all the available decadal predictions contributing to the Decadal Climate Prediction Project (DCPP; Boer et al., 2016) of the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016). Information on such decadal predictions and the forecast system can be seen in

Table 2.

Table 2: Forecast systems contributing to the DCPP-A component of the CMIP6 and their specifications (available simulations at the time of the study). The spatial resolution is shown for the atmospheric grid as 'nlon x nlat', where nlon corresponds to the number of longitudes and nlat to the number of latitudes of the respective global grid.

Forecast system	Institution	nº of DCPP members	nº of HIST members	Spatial resolution	Month of initialization	Reference
BCC-CSM2-MR	BCC	8	3	320 x 160	January	Wu et al. (2019)
CanESM5	CCCma	20	40	128 x 64	January	Swart et al. (2019)
CESM1-1-CAM5-CMIP5	NCAR	40	40	288 x 192	November	Yeager et al. (2018)
CMCC-CM2-SR5	CMCC	10	1	288 x 192	November	Cherchi et al. (2018)
EC-Earth3-i1	BSC	10	10	512 x 256	November	Bilbao et al. (2021)
EC-Earth3-i2	SMHI/DMI	5	-	512 x 256	November	Tian et al. (2021)
HadGEM3-GC3.1-MM	MOHC	10	4	432 x 324	November	Sellar et al. (2020)
IPSL-CM6A-LR	IPSL	10	32	144 x 143	January	Boucher et al. (2020)
MIROC6	MIROC	10	10	256 x 128	November	Tatebe et al. (2019)
MPI-ESM1.2-HR	DWD	10	10	384 x 192	November	Müller et al. (2018)
MPI-ESM1.2-LR	DWD	16	10	192 x 96	November	Mauritsen et al. (2019)
MRI-ESM2-0	MRI	10	5	320 x 160	November	Yukimoto et al. (2019)
NorCPM1	NCC	10	30	144 x 96	October	Bethke et al. (2021)

The evaluation has been performed over the Southern African Development Community (SADC) during the 1966-2014 period with the anomalies of the considered variables to remove the mean bias of the predictions. The 1981-2010 period has been used to compute the climatology and thresholds between the probabilistic categories. Each year, the predictions are run for the next ten years.

Different forecast periods have been considered (forecast years 1, 1-5, 6-10 and 1-10). Different reference datasets (gridded observations and reanalyses) have been used for each variable to account for the observational uncertainty (Table 3).

Table 3: Reference datasets used for the forecast quality assessment and their specifications. The spatial resolution is shown for the atmospheric grid as 'nlon x nlat', where nlon corresponds to the number of longitudes and nlat to the number of latitudes of the respective global grid.

Variable	Reference dataset	Institution	Type	Spatial resolution	Reference
Near-surface air temperature	GHCNv4	NOAA	Gridded observations	72 x 36	Menne et al. (2018)
	JRA-55	JMA	Reanalysis	288 x 145	Kobayashi et al. (2015)
	ERA5	ECMWF	Reanalysis	1280 x 640	Hersbach et al. (2020)
Sea level pressure	JRA-55	JMA	Reanalysis	288 x 145	Kobayashi et al. (2015)
	HadSLP2	MOHC	Gridded observations	72 x 37	Allan and Ansell (2006)
	ERA5	ECMWF	Reanalysis	1280 x 640	Hersbach et al. (2020)
Precipitation	GPCC	DWD	Gridded observations	360 x 180	Schneider et al. (2018)
	JRA-55	JMA	Reanalysis	288 x 145	Kobayashi et al. (2015)
	ERA5	ECMWF	Reanalysis	1280 x 640	Hersbach et al. (2020)

Verification is an essential step to estimate the quality and reliability of the forecasts. For it, retrospective decadal predictions (known as decadal hindcasts) are created from 1960 to have a large enough sample size to obtain robust quality estimates by comparing them against past observations. The hindcasts are run with the same forecast systems as the actual forecasts and allow the detection of systematic errors in the forecasts that can be partially corrected by applying post-processing techniques such as bias-adjustment and calibration (Doblas-Reyes et al., 2005).

In this work five different calibration techniques have been applied with the aim of correcting the main types of model's biases. Besides, the quality of the calibrated predictions has been compared to that of the raw predictions to assess the impact of calibration. The five calibration methods are the following: ‘bias’ (which corrects only the mean bias; Torralba et al., 2017), 'evmos' (which applies a

variance inflation technique to ensure the correction of the mean bias and the correspondence of the variance between the predictions and observations; Van Schaeybroeck and Vannitsem, 2011), 'mse_min' (which corrects the mean bias, the overall forecast variance and the ensemble spread by minimising a constrained mean-squared error; Doblas-Reyes et al., 2005), 'crps_min' (which corrects the mean bias, the overall forecast variance and the ensemble spread while minimising the continuous ranked probability score; Van Schaeybroeck and Vannitsem, 2015) and 'rpc_based' (which adjusts the forecast variance ensuring that the ratio of predictable components is equal to one; Eade et al. 2014). All the calibration methods are implemented in the CSTools R-package (Pérez-Zanón et al., 2022) and have the option to be applied in leave-one-out cross-validation mode (i.e., without using information of the time step that is being calibrated, as it would be done in a real-time context).

Several metrics have been considered to assess different aspects of the quality of both deterministic and probabilistic forecast products. Such evaluation has been systematically applied to all the forecast systems, variables and forecast periods. The Anomaly Correlation Coefficient (ACC; Wilks, 2011) and the Root Mean Squared Error Skill Score (RMSSS; Wilks, 2011) have been used for the deterministic products. For the probabilistic products based on tercile and quintile equiprobable categories, the Ranked Probability Skill Score (RPSS; Wilks, 2011), the FairRPSS (Ferro, 2014) and the Relative Operating Characteristic Skill Score (ROCSS; Kharin and Zwiers, 2003) have been used. Those metrics ending with the suffix 'Skill Score' allow comparing the performance of the predictions against other reference forecasts (e.g., climatology, persistence, or other predictions previously used by users; Jolliffe and Stephenson, 2012). In this work, the climatological forecast (defined as the equiprobable forecast, i.e., probability of 33.33% for each tercile category) has been used as the reference forecast. The skill scores range between minus infinite and 1. If a skill score is greater than 0, it indicates that the forecast has higher skill than the reference forecast, while a negative value means that it has a lower skill. Besides, the Spread-over-Error ratio (Fortin et al., 2014), Signal-to-Noise ratio (Scaife and Smith, 2018), and Ratio of Predictable Components (RPC; Eade et al., 2014) have also been calculated. A two-sided t-test has been applied to estimate the ACC significance at the 95% confidence level accounting for the time series' autocorrelation (Von Storch and Zwiers, 2001). The Random Walk test (DelSole and Tippett, 2016) has been used to assess the significance of the skill scores at the 95% confidence level.

Given the large number of results, only some have been shown here: the ones corresponding to the forecast quality measured with the RMSSS and RPSS for tercile categories for near-surface air temperature predictions for the forecast years 1-5 using the ERA5 reanalysis as the reference dataset. The rest of the results can be accessed through a R Shiny App (https://earth.bsc.es/shiny/FOCUS-Africa/) created to facilitate the accessibility and visualisation of the figures.

3.2 Results

The quality of the deterministic forecasts of near-surface air temperature measured with the RMSSS shows a benefit of using the decadal predictions compared to the climatological forecast for most regions considered (as indicated by positive values in Figure 50). Still, the regions where the RMSSS is statistically significant vary across the forecast systems, finding most of them over Madagascar and the central and western parts of the SADC. In contrast, there are regions where the climatological

forecast shows a higher quality than the decadal predictions (e.g., over Botswana for some forecast systems), although they are not statistically significant.

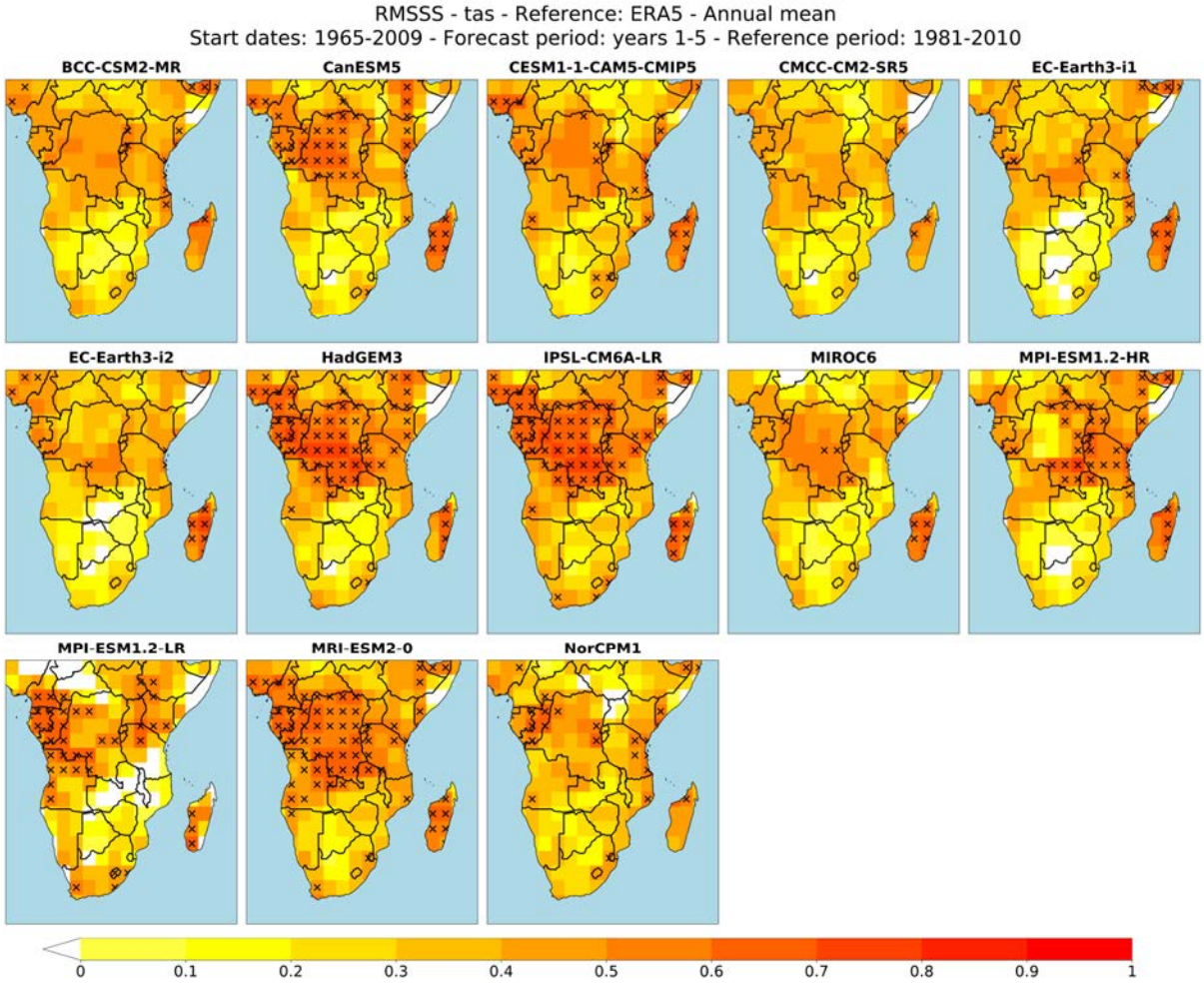


Figure 50: RMSSS of the CMIP6/DCPP forecast systems for the near-surface air temperature for the forecast years 1-5 using climatology as the reference forecast. The predictions have been evaluated over the 1966-2014 period (start dates 1965-2009) for each grid point. The reference period to compute the anomalies is 1981-2010. The reference dataset is the ERA5 reanalysis. Crosses indicate that the values are statistically significant at the 95% confidence level based on a Random Walk test.

The systematic evaluation applied to all the forecast systems allows detecting windows of opportunity (e.g., regions, forecast periods and variables) with enough skill to provide a climate service. For instance, the RMSSS values can be useful for selecting the forecast system that provides the highest forecast quality for each grid point. In that case, a different forecast system can be selected for each region to ensure the highest possible quality and most reliable predictions for the forthcoming years. Also, another advantage of such selection (in comparison to producing a forecast using only one model) is that a particular forecast system might be skilful over a region, but it might not be over others. Thus, the forecast system that provides the highest RMSSS for each grid point has been selected and shown in Figure 51. Besides, the percentage of grid points where each system is the most skilful is also shown. For this case (forecasts of near-surface air temperature for the forecast years 1-

5), the IPSL-CM6A-LR is the forecast system that provides the highest skill (measured with the RSMSS) over more grid points of the SADC region (29.15% of the region).

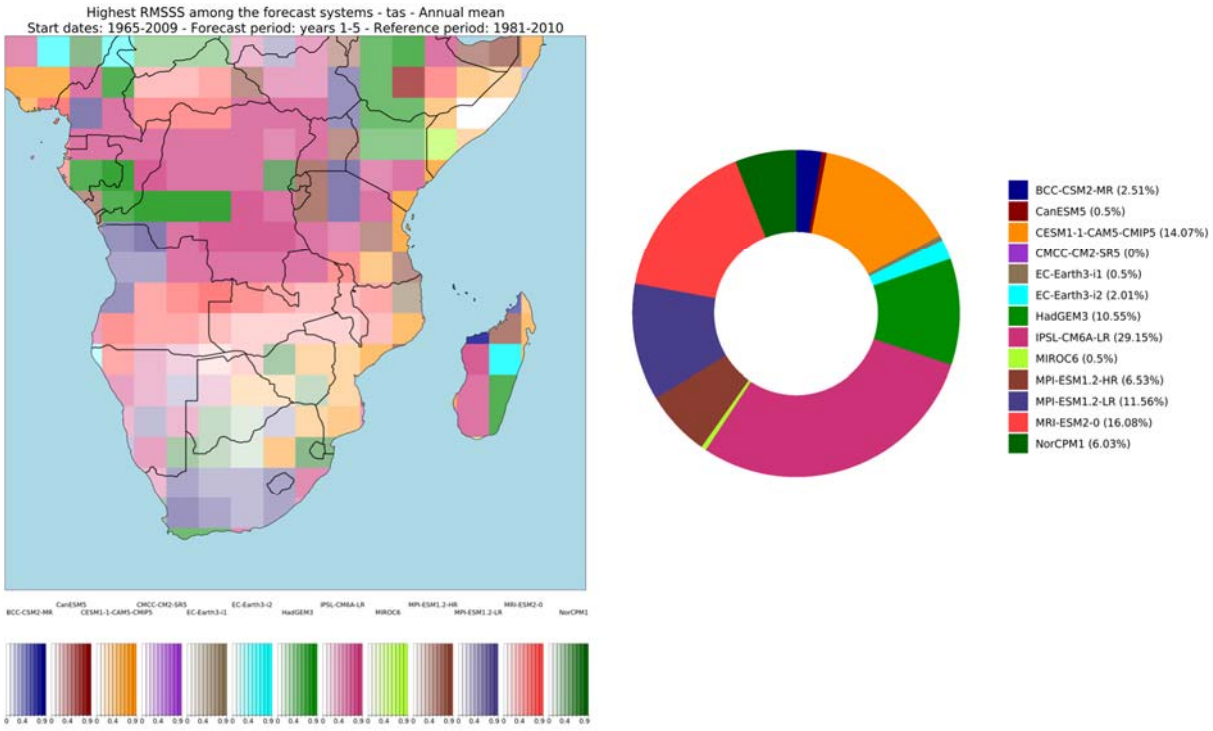


Figure 51: Highest RMSSS of the CMIP6/DCPP forecast systems for the near-surface air temperature for the forecast years 1-5 using climatology as the reference forecast (left) and Doughnut chart with the percentage of grid points where each forecast system provides the highest RMSSS (right). The predictions have been evaluated over the 1966-2014 period (start dates 1965-2009) for each grid point. The reference period to compute the anomalies is 1981-2010. The reference dataset is the ERA5 reanalysis.

Similar to the quality assessment of the deterministic products, the probabilistic forecasts have been evaluated (Figure 52). The RPSS for tercile categories shows an added value of the decadal predictions compared to the climatological forecast (defined as equal probabilities for all the tercile categories, i.e., 33.33% of probability of occurrence for the three probabilistic categories). Besides, the RPSS values are statistically significant for most regions and forecast systems. On the other hand, the climatological forecast provides higher skill than the decadal predictions over some regions like Botswana and its surroundings. However, these negative RPSS values are not statistically significant.

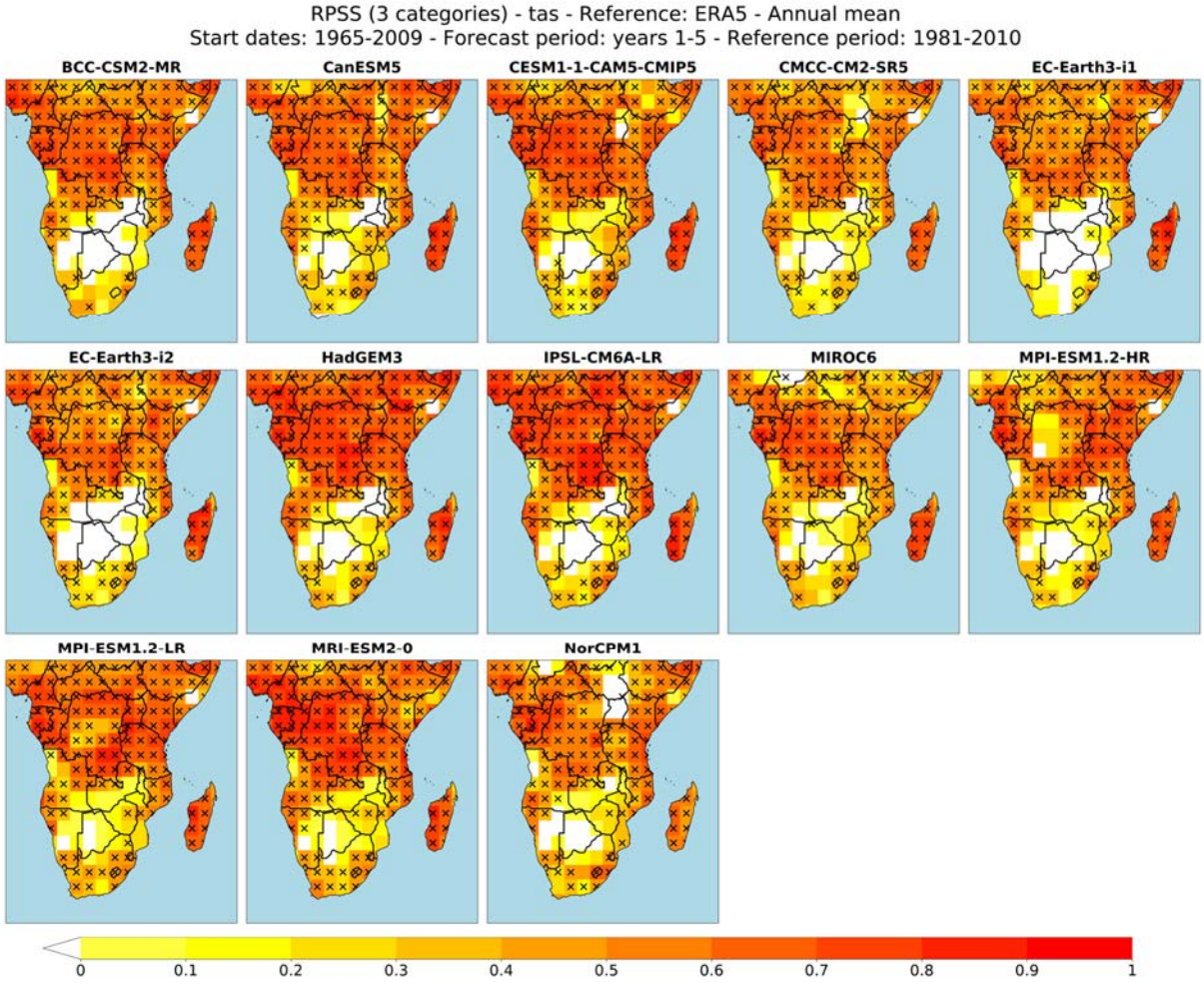


Figure 52: RPSS for tercile categories of the CMIP6/DCPP forecast systems for the near-surface air temperature for the forecast years 1-5 using climatology as the reference forecast. The predictions have been evaluated over the 1966-2014 period (start dates 1965-2009) for each grid point. The reference period to compute the tercile categories is 1981-2010. The reference dataset is the ERA5 reanalysis. Crosses indicate that the values are statistically significant at the 95% confidence level based on a Random Walk test.

In addition, there are other regions where some forecast systems do not provide better probabilistic predictions than climatology, for example, over South Sudan and Uganda with the NorCPM1 system. However, the possibility to select the best forecast system for each specific region could provide the highest quality forecast for all the regions (Figure 53). For this specific probabilistic forecast evaluation, three models stand out for showing the highest skill over a large percentage of the region. These are the MRI-ESM2-0 (which is the best system over the 23.12% of the region), IPSL-CM6A-LR (20.6%) and HadGEM3-GC3.1-MM (19.1%) forecast systems.

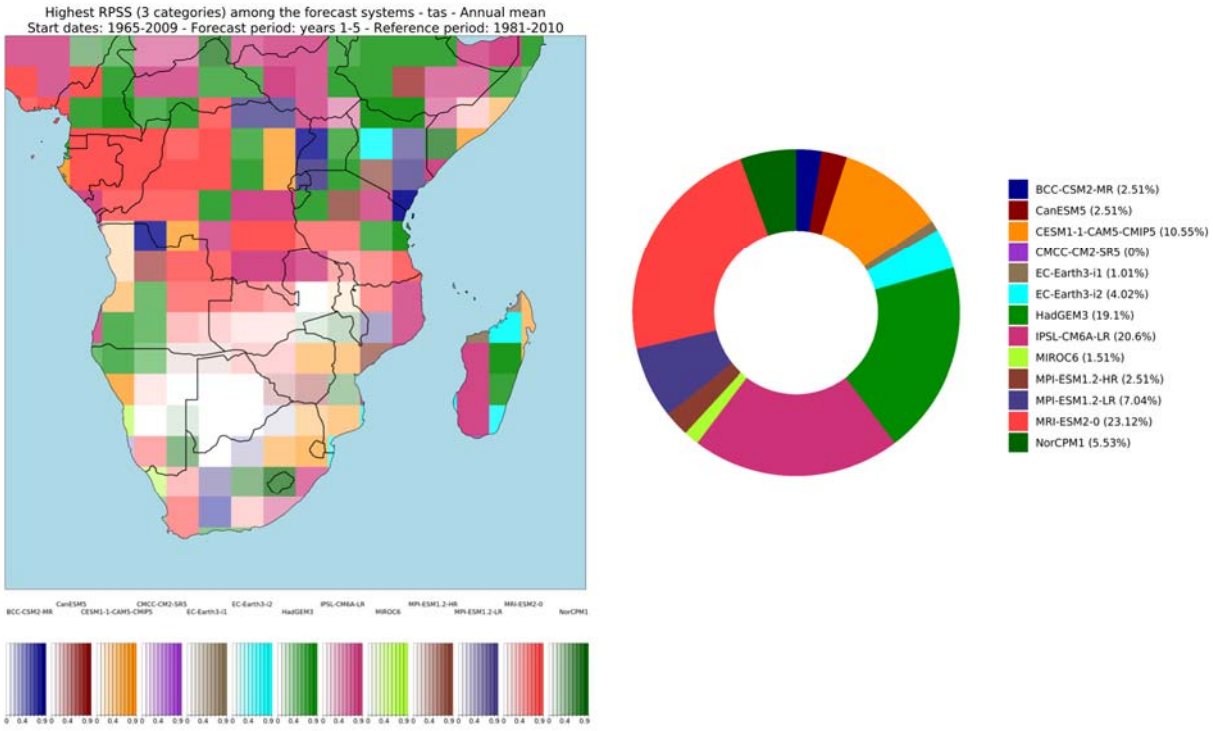


Figure 53: Highest RPSS for tercile categories of the CMIP6/DCPP forecast systems for the near-surface air temperature for the forecast years 1-5 using climatology as the reference forecast (left) and Doughnut chart with the percentage of grid points where each forecast system provides the highest RPSS (right). The predictions have been evaluated over the 1966-2014 period (start dates 1965-2009) for each grid point. The reference period to compute the tercile categories is 1981-2010. The reference dataset is the ERA5 reanalysis.

As it has been introduced in section 3.1, five different calibration methods have been applied to the raw decadal predictions to improve the forecast products' quality and reliability. For each forecast system, the calibrated predictions have been evaluated and compared against the uncalibrated predictions to estimate the impact of these calibration techniques. The calibration has been applied in leave-one-out cross-validation mode, i.e., without using the information of the time step that is being calibrated in order to not overestimate the impact of calibration.

For the deterministic temperature forecast for the forthcoming five years produced with the EC-Earth3-i1 system, the RMSSS values obtained with the raw and calibrated decadal predictions show an added value of using such predictions in comparison to the climatological forecast over the SADC region (Figure 54). The raw forecasts show positive values over most parts of the region considered. However, the RMSSS values are not statistically significant, except for some grid points in the eastern sector. Instead, the calibrated forecasts show more regions with significantly positive RMSSS values, indicating a benefit of applying calibration techniques, particularly over the central and northern parts of the SADC region. Such benefit is also shown when comparing the calibrated and raw predictions, as the RMSSS values are generally positive (and significant over several regions).

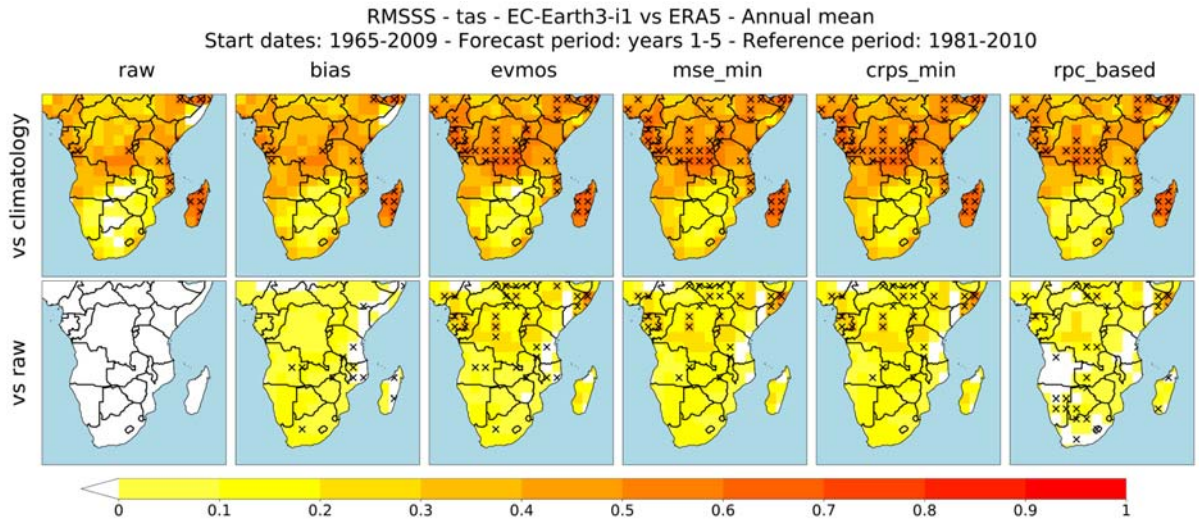


Figure 54: RMSSS of the EC-Earth3-i1 forecast system for the near-surface air temperature for the forecast years 1-5 for the raw and calibrated predictions using climatology as the reference forecast (top row) and for the calibrated predictions using the raw predictions as the reference forecast (bottom row). The predictions have been evaluated over the 1966-2014 period (start dates 1965-2009) for each grid point. The reference period to compute the anomalies is 1981-2010. The reference dataset is the ERA5 reanalysis. Crosses indicate that the values are statistically significant at the 95% confidence level based on a Random Walk test.

The predictions might have different bias types and magnitude over different locations. Thus, the calibration method that most improves the forecast quality might differ across regions. Also, it should be noted that each calibration approach aims at correcting only some of the model's biases and, thus, they can only improve specific aspects of the forecast quality as measured with each specific metric. Therefore, the forecast quality might decrease if measured with another metric, and the calibration method thus should be chosen accounting for the aspect of the quality that needs to be improved (which strongly depends on the specific user's needs). For the case of the RMSSS (Figure 55), the calibration methods that show the highest benefit are the 'evmos' (which is the best one over 32.66% of the region), 'mse_min' (25.13%), and 'bias' (18.59%) methods.

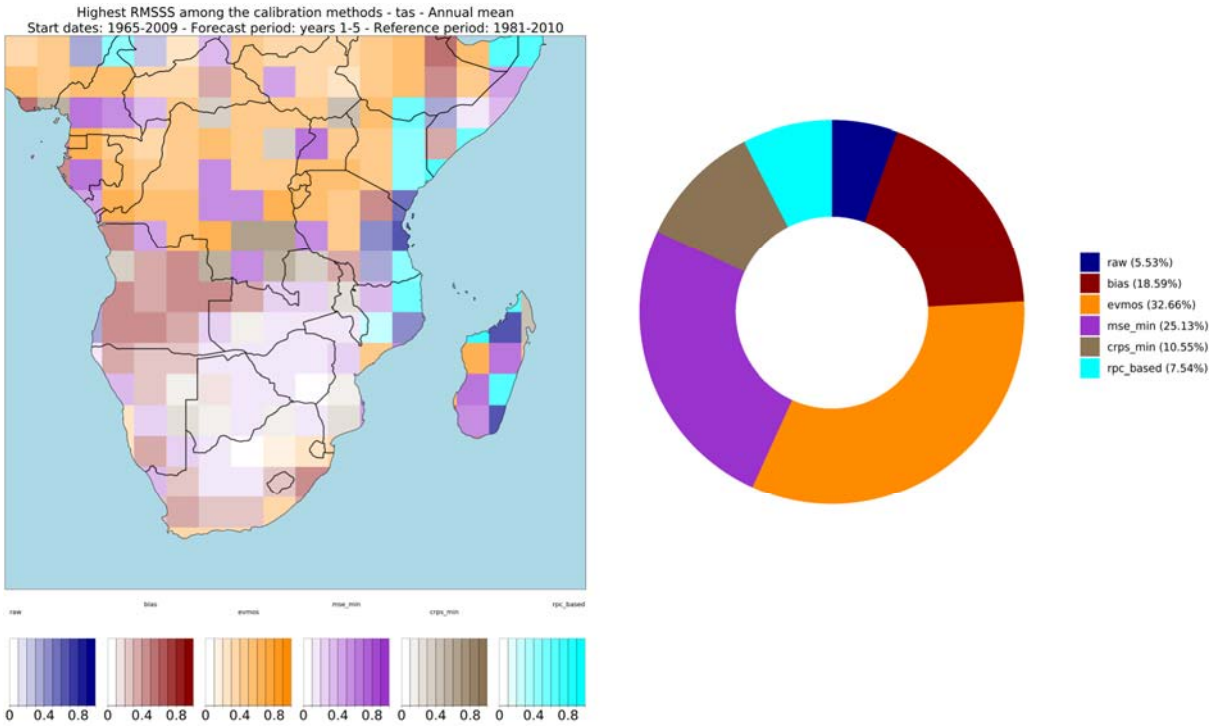


Figure 55: Highest RMSSS of the raw and calibrated predictions for the near-surface air temperature for the forecast years 1-5 using climatology as the reference forecast (left) and Doughnut chart with the percentage of grid points where each calibration method provides the highest RMSSS (right). The predictions have been evaluated over the 1966-2014 period (start dates 1965-2009) for each grid point. The reference period to compute the anomalies is 1981-2010. The reference dataset is the ERA5 reanalysis.

The probabilistic raw and calibrated predictions have also been compared. When measured with the RPSS, the calibrated predictions generally show little or no quality improvements with respect to the raw predictions (Figure 56). The sector that shows the highest benefit from calibration is the Southern part of the SADC region. However, it mostly coincides with the sector where neither the raw predictions nor the calibrated ones show skill.

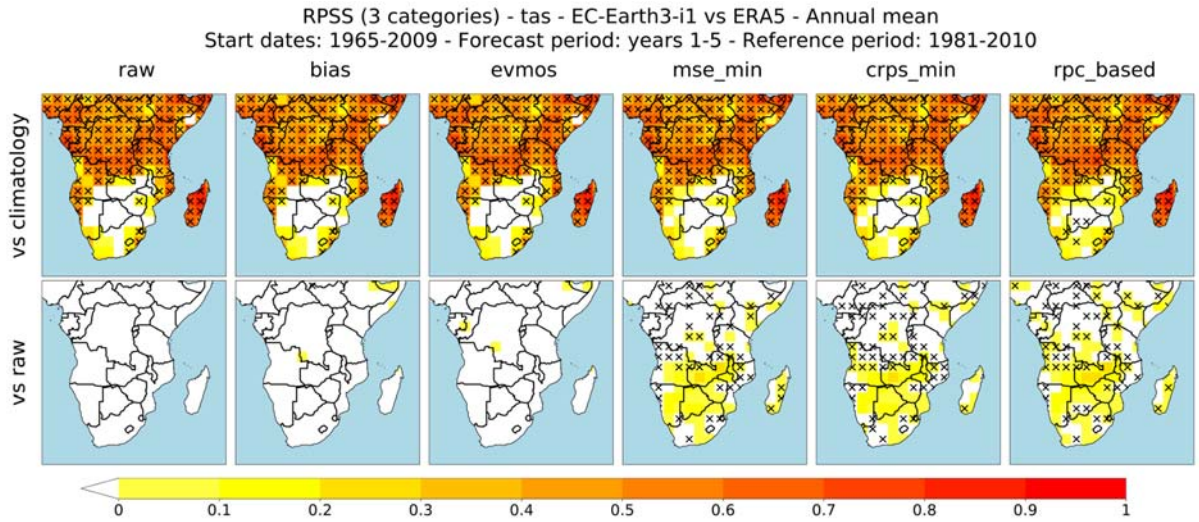


Figure 56: RPSS for tercile categories of the EC-Earth3-i1 forecast system for the near-surface air temperature for the forecast years 1-5 for the raw and calibrated predictions using climatology as the reference forecast (top

row) and for the calibrated predictions using the raw predictions as the reference forecast (bottom row). The predictions have been evaluated over the 1966-2014 period (start dates 1965-2009) for each grid point. The reference period to compute the tercile categories is 1981-2010. The reference dataset is the ERA5 reanalysis. Crosses indicate that the values are statistically significant at the 95% confidence level based on a Random Walk test.

The low improvement due to calibration found when using the RPSS might be a consequence of this specific skill score not being sensitive to biases in the mean and variance (contrary to, for example, the RMSSS). Then, although the mean and variance are adjusted (among other statistical properties), the RPSS does not account for it. In this case, the calibrated forecast with the 'rpc-based' calibration method is the one that shows the highest forecast quality over most points (34.17% of the considered region; Figure 57).

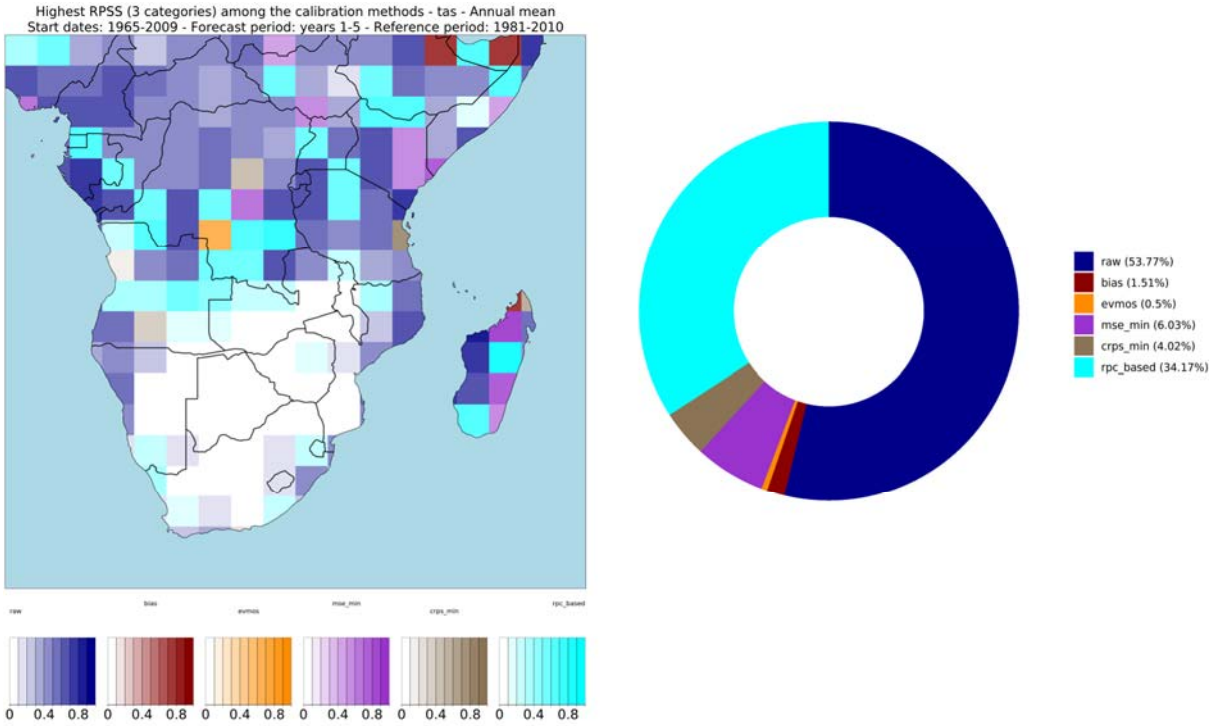


Figure 57: Highest RPSS for tercile categories of the raw and calibrated predictions for the near-surface air temperature for the forecast years 1-5 using climatology as the reference forecast (left) and Doughnut chart with the percentage of grid points where each calibration method provides the highest RPSS (right). The predictions have been evaluated over the 1966-2014 period (start dates 1965-2009) for each grid point. The reference period to compute the tercile categories is 1981-2010. The reference dataset is the ERA5 reanalysis.

4 Climate projections

The southern African region was classified as a climate change hotspot by the Intergovernmental Panel on Climate Change (IPCC) Special Report on Global Warming of 1.5 °C (Hoegh-Guldberg et al., 2018). This stems from the region being projected to become drastically warmer, and likely also drier (Engelbrecht et al., 2015; Hoegh-Gulberg et al., 2018; Lee et al., 2021). In fact, the IPCC has recently assessed that general trends of drying and substantial warming can already be detected across the region. In eastern southern Africa, including the eastern escarpment region of South Africa and Mozambique, climate models are projecting increases in intense rainfall events, despite the general reduction in rainfall totals (Ranasinghe et al., 2021). This aspect of the projected climate change signal can also already be detected in observed statistics of intense rainfall events occurring in eastern southern Africa over the last few decades (Ranasinghe et al., 2021). In Mozambique increases in the intensity of tropical cyclones (Fitchett, 2018), or at least in the rainfall they produce, likely play a role in the upward trend in the number of recorded intense rainfall events. A recent climate change attribution study has consistently assessed that climate change has likely resulted in an increase in precipitation associated with the series of tropical cyclones that made landfall in Mozambique in 2022 (Otto et al., 2022).

Average temperatures have been increasing at about twice the global rate of temperature increase over the southern African interior, with highest rates of temperature increased recorded over northern Botswana and southern Zambia (Engelbrecht et al., 2015; Kruger and Nxumalo, 2016). Extreme temperature events such as very hot days, heat-wave days and high fire-danger days have correspondingly increased drastically in their frequency of occurrence over the last several decades (Kruger and Sekele, 2013). It is certain that further increases in oppressive temperature events will occur in the region for as long as global warming continues (Garland et al., 2015; Seneviratne et al., 2021).

4.1 Characterization of climate projections for temperature and rainfall

4.1.1 Summary of findings from D3.1

GCMs are widely used in the assessment of current and future climates on global and regional scales. Due to current knowledge on processes affecting the climate system on centennial timescales, models are constructed from basic components (e.g., atmosphere, ocean, cryosphere, land surface). Since not all these processes can be fully described from basic principles and fully resolved at GCM spatial scales, several assumptions are needed in building GCMs. GCMs also need assumptions on external forcings, e.g., anthropogenic greenhouse gas emissions and aerosols, resulting in a varied range of climate simulations. D3.1 considered the more recent CMIP6 ensemble with an increased spatial resolution over its predecessor (CMIP5), allowing for a better representation of smaller scale physical processes (Eyring et al., 2016; O'Neill et al., 2017; Stouffer et al., 2017) as well as additional forcings which include the Shared Socio-Economic Pathway (SSP) scenario matrix (Eyring et al., 2016; Riahi et al., 2017; Stouffer et al., 2017). The main findings from this deliverable are the following:

- **Temperature:**
  - There is a cold bias across most of SADC during the warm season (DJF), with the west coast being the only exception to this.
  - JJA has a reduced cold bias and more variable results. These results could suggest a muted seasonal cycle across CMIP6.
- **Precipitation:**
  - Wet seasons rainfall is varied across all models, but there tends to be more of a wet bias.
  - The bias is smaller in the drier months (JJA), but with a systematic bias over South Africa.

The same analysis was conducted for future projections under RCP8.5:

- **Temperature:**
  - Increase in mean temperature shown across all models ranging from 3 ^{o}C to 9 ^{o}C.
  - Greatest increase in the southern and western interior of SADC.
  - CMIP6 shows a stronger warming in the 95th percentiles of the ensemble range in comparison to the CMIP5, CORDEX-AFR44 and CORDEX-AFR22.
- **Precipitation** (see Table 4 for summary):
  - CMIP6 shows little consistency across the ensemble members, a result which is similar to its predecessor (CMIP5)
  - For most case study regions, there is no statistically significant increase or decrease in mean precipitation over the December, January, February season. However, several models agree that Tanzania could see a significant increase in rainfall.
  - For the June to September season, there is still no significant change across most models. However, a handful of models do show a significant drying across south and eastern SADC, which would impact several case study countries.
  -

Table 4: summary of the direction of projected changes in rainfall by the end of the century across the regions studied in D3.1. Where "+" indicates and increase "-" indicates a reduction and "=" indicates no significant change. Th green boxes show areas with significant discrepancies between models.

Precipitation	LOWER ESAF				LOWER WSAF				UPPER ESAF				SEAF			
MODEL	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
CMIP6	+	-	-	-	-	-	-	-	+	+	-	-	+	+	=	+

4.2 Characterization of climate projections in simulating climate extremes

4.2.1 Global Circulation Models (GCMs)

4.2.1.1 Performance over case study countries

Here an overview is provided of the findings from two studies looking at TXx (Almazroui, Saeed, et al., 2021) and RX1 (Dosio et al., 2021) for the CMIP6 ensemble over Africa to compare their performance to that of RCMs. TXx was investigated by Almazroui et al. (2021), looking at spatial representations and future projections of CMIP6 over the IPCC's Special Report for Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX) regions. One finding from this study was the marked similarity between the spatial representation of TXx between CMIP6 and the CMIP5 ensemble (when compared to the study by Seneviratne and Hauser, 2020).

Dosio et al. (2021) conduct an evaluation of CMIP5 & 6, CORDEX and CORDEX-CORE future daily precipitation over Africa. Like the TXx studies, there were no clear improvements between CMIP5 & 6 ensembles for this variable. RX1 is generally underestimated by CMIP5 and CMIP6, while CORDEX datasets are showing a better agreement over the continent, except for Southern Africa, where CORDEX have a positive bias.

When looking at the CMIP6 ensemble performance over Africa, the ensemble projects the greatest increase in TXx over South Africa, with the tropical regions seeing the smallest increase. This was seen globally, with extratropical regions projecting a greater increase in TXx than tropical regions. Future projections of mean precipitation across the continent from this study found robust increases over 10-20 % (land coverage) of Southern Africa and ~35 % of East Africa. The remaining land coverage has little agreement across the ensemble, highlighting the uncertainty of future rainfall in this region, which is captured better in a larger ensemble. However, maximum daily intensity is projected to increase in most models.

4.2.1.2 Limitations

Table 5: Limitations of the CMIP6 GCM models.

Limitation	*Impact*
Coarse resolution	The coarse resolution of the GCM models impact their ability to recreate mesoscale characteristics of rainfall that impact intensity. This is where RCMs add value, helping in the simulation of higher order precipitation statistics and characteristics. This has been noted in a number of studies (e.g. Dosio et al. 2015; Pinto et al. 2016; Nikiema et al. 2017; Fotso-Nguemo et al. 2017; Gibba et al. 2019; Tamoffo et al. 2020; Gnitou et al. 2021)

4.2.2 Regional Climate Model Datasets (RCMs)

4.2.2.1 CP4A

4.2.2.1.1 Performance over case study countries

The main incentive to use CP4A comes from its ability to better represent convection. It has been shown that these *convection permitting* models are able to capture rainfall characteristics at an hourly timescale (Stratton et al., 2018), which could be of use to some sectors. Conversely, RCMs with implicit convection and GCMs (Nikulin et al., 2012) show deficiencies in the representation of the diurnal cycle of precipitation. However, D3.3 found that there was no improvement to average rainfall, which has been noted in another study (Finney et al., 2019). Additionally, its representation of RX1 was far wetter than observations similar to the findings of Kendon et al. (2019), although still able to represent very extreme hourly events. In our analysis, most case study countries had a bias of 50 – 60+ mm day^{-1}, showing a need for bias correction before the data can be used.

4.2.2.1.2 Limitations

The evaluation of CP4A was limited to the historical period in D3.3 due to '*the limited 10-year time slice making its application very limited across the case studies'.* This was reason enough to exclude it from further investigation; however, for full transparency; here an overview of additional limitations of this dataset is also provided (Table 6) .

Table 6: Limitations of the CP4A model as stated in (Senior et al., 2020)

Limitation	Impact
Single ensemble member and driving model	With one driving model, cannot capture uncertainty in the simulation of global climate and how this impacts the Africa climate. Additionally, having a single ensemble member makes it difficult to infer the significance of the results.
Single time slice available for future projections	Does not allow for mid-century projections and will limit the analysis on statistics of annual extremes.
All soil has the same properties	In this model, all soil in the land component of the model is defined as sandy.While a good agreement has been found on temperature indices, including extremes, it is not possible to infer that this approximation will also be effective in the future climate. In addition, impacts dependent on soil moisture and radiation will not be feasible since these variables are not physically representative of the grid box properties.

Despite these limitations there may be some use cases for this model for hourly extremes, for instance, in hydrological modelling and catchment case studies. There is also the potential for sub-daily analysis using CP4A, but the setup is restricted, only being able to represent the end of the century under RCP8.5, and a GCM with a high climate sensitivity. Therefore, it is suggested that using

this model on its own will not provide a true representation of future uncertainties and other models should accompany it in any case study analysis.

4.2.2.2 CORDEX

4.2.2.2.1 Performance over case study countries

In addition to the analysis performed in D3.3, a country specific analysis of the short duration's events identified in D3.3 (TXx, TNn and RX1) is also provided. This overview combined with the analysis already conducted allows us to provide a good overview of this dataset and its limitations over case study regions.
Figure **58** to Figure 62 provide an average TXx, TNn and RX1 value for each case study country, compare this to observations (Historical) and provide the climate change signal, using the RCP8.5 scenario. It is worth noting that due to the size of Mauritius, analysis was not feasible over this country and has therefore not been included in the following discussion.

Figure **58**Figure 62 show the average TXx, TNn and RX1 observational biases (left hand columns) and projected changes under RCP8.5 (right hand columns), over the case study countries. Here, temperature biases and projections show change in ^{o}C, whereas RX1 shows bias from observations in mm /day and future projections as percentage change from the historical climatology.

Tanzania

From D3.3 it can be seen the representation of TXx from the high-resolution model was good over Tanzania. This is backed up by the averages taken over the country, with biases within +/- 1.5^{o}C and an ensemble average of 0.3 ^{o}C (
Figure **58**). In the coarser resolution ensemble, there are slightly more varied results, with a colder bias and averages ranging from 0.2 ^{o}C to -4.0 ^{o}C. TNn is not captured as well, with all models across all resolutions underestimating TNn values. Here, biases range from 0.5 – 7.2 ^{o}C and both ensemble averages are > 4 ^{o}C.

RX1 in the CORDEX AFR-22 ensemble has a very clear wet bias, visible in both the analysis in D3.3 and in Figure 1. A strongest bias can be seen in the higher resolution CCCMA-CANRCM4 (+49.21 mm day^{-1}). The CORDEX AFR-44 ensemble, again, has a far more varied output, with strong dry (RACMO22T = > -30 mm day^{-1}) and wet biases (CCLM4_8_17 up to 48 mm day^{-1}). However, it improves on the higher resolution dataset when looking at the ensemble averages, possibly because of the larger size of the CORDEX AFR-44 ensemble.

Projections for Tanzania see TXx and TNn values increasing for both variables by about 3 -5 ^{o}C. RX1 is also projected to increase in all models over Tanzania, with ensemble averages suggesting a 28.51 % increase in CORDEX AFR 22 and a 26.59% increase in CORDEX AFR-44.

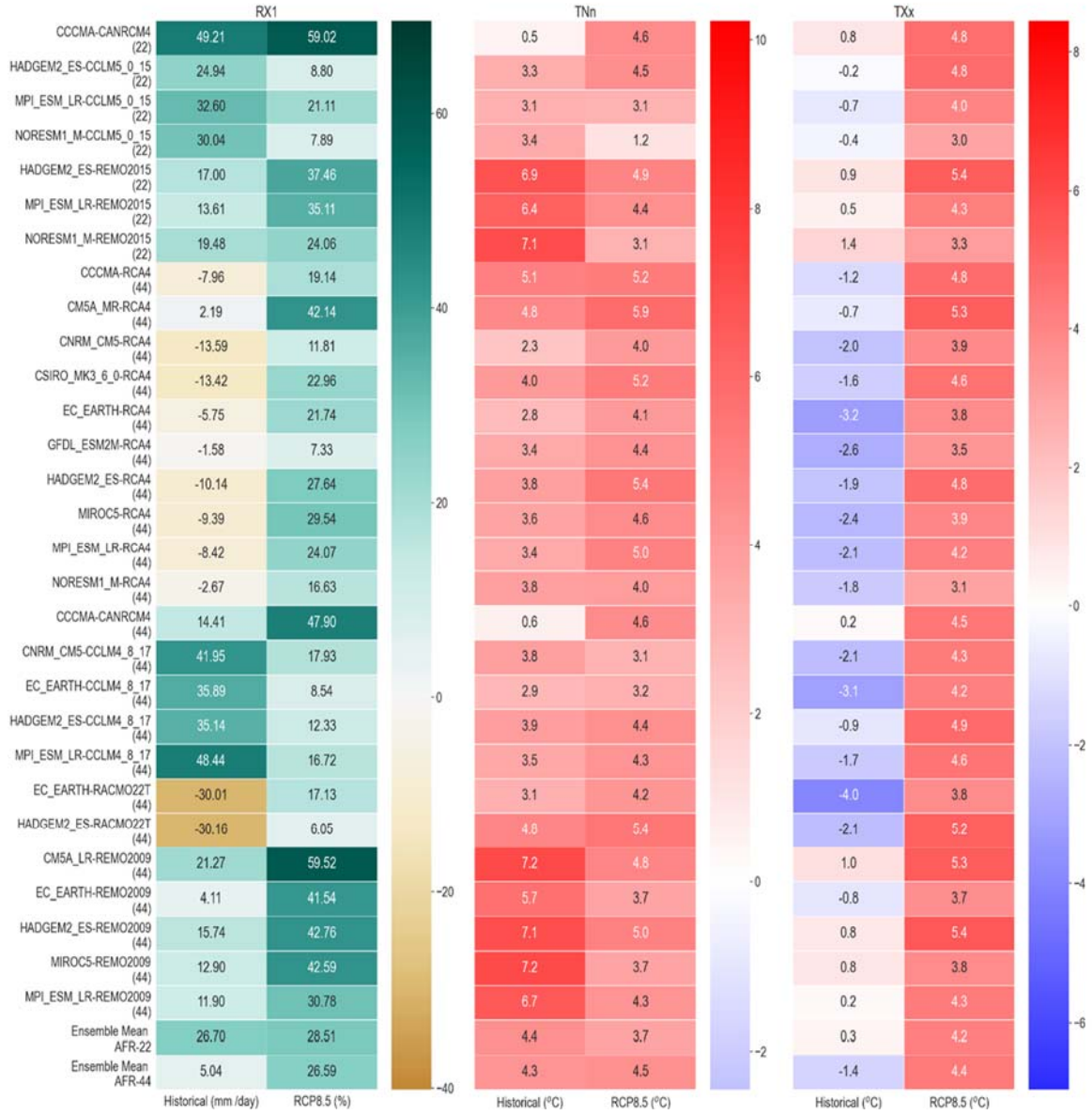


Figure 58: Averages for extreme variables (TXx, TNn and RX1) over Tanzania. With Historical representing their bias compared to observations (CPC for TXx and TNn, and CHIRPS for RX1) and RCP8.5 comparing their end of century climate projection (2070 – 2100) to their historical climatology (1975 – 2005).

Malawi

TXx over Malawi is well represented by these models at both resolutions (Figure 59). The higher resolution CORDEX ensembles do tend to slightly overestimate TXx values here, with biases up to 2.6 °C (CCCMA-CANRCM4). However, the MPI_ESM_LR driven models fall within 0.2 °C of observations over Malawi. The CORDEX AFR-44 ensembles have mixed results, with RACMO22T and RCA4 generally having a colder bias and REMO2009 having a noticeable warmer bias. TNn again is underestimated by most models apart from CCCMA-CANRCM4 (> -3°C) and some RCA4 runs (< -1 °C). Again, REMO models have a notably warmer bias regardless of driving models. However, the general performance across the ensembles is good with CORDEX AFR-22 having a bias of 2.8 °C and CORDEX AFR-44 being 1.9°C.

RX1 is overestimated considerably by the higher resolution dataset, with an ensemble average of > 50 mm day^{-1}. The CORDEX AFR-44 models perform slightly better than their higher resolution counterparts, with an ensemble average of 24.61 mm day $^{-1}$. The RCA4 models are closest to observed values, with most runs being within 10 – 15 mm day^{-1}. Regarding projections, they indicate again a rise in TXx and TNn values of between 2 – 6 °C and RX1 values also project an ensemble average increase of just over 17% at both resolutions.

Figure 59: Averages for extreme variables (TXx, TNn and RX1) over Malawi. With Historical representing their bias compared to observations (CPC for TXx and TNn, and CHIRPS for RX1) and RCP8.5 comparing their end of century climate projection (2070 – 2100) to their historical climatology (1975 – 2005).

South Africa

South Africa had one of the more consistent spatial representations in D3.3 with most models underestimating TXx, particularly around the high elevations surrounding Lesotho. The high-resolution model ensemble has a more diverse output with CCLM5_0_15 and CANRCM4 downscaled simulations having a warm bias (0 – 2.5 °C) and the REMO simulations having a colder bias (1.5 – 2.4 °C). This variance results in an ensemble average within 0.1 °C of observations. The coarser resolution ensemble has a clear colder bias here, the CCLM4_8_17 models are closest to observations (+/- 1°C), but most downscaled simulations underestimate TXx by ~2 °C to 4 °C. For TNn, the CANRCM4 models at both resolutions are the only models to significantly overestimate this variable (> 3 °C). The

remaining models underestimate TNn with REMO being 4 °C to 5 °C warmer than observations. The RACMO22T simulations show the best agreement over this region from both D3.3 and Figure 60 with values close to observations.

RX1 is overestimated by all CORDEX AFR-22 models, with NORESM1_M – REMO2015 member being farthest from the observed value (+ 38.52 mm day^{-1}). The CORDEX AFR-44 models are generally closer to observations with RCA4 downscaled simulations performing well. There is still a wetter bias in the REMO and CCLM4_8_17 downscaled models comparatively. TXx and TNn still show the standard increases, with ensemble average increases of 5.1°C and 3.1 / 3.2 °C respectively. However, projections for this country are interesting when it comes to precipitation. Both resolutions have mixed results, with ensemble averages indicating a slight increase in RX1 values. However, when looking at models individually RCA4 and CCLM (both resolutions) showed little change in RX1 values, ranging between +/- 5mm day $^{-1}$. REMO generally projects the greatest increase compared to the historical climatology (up to 19 mm day $^{-1}$), although this is still small compared to other countries.

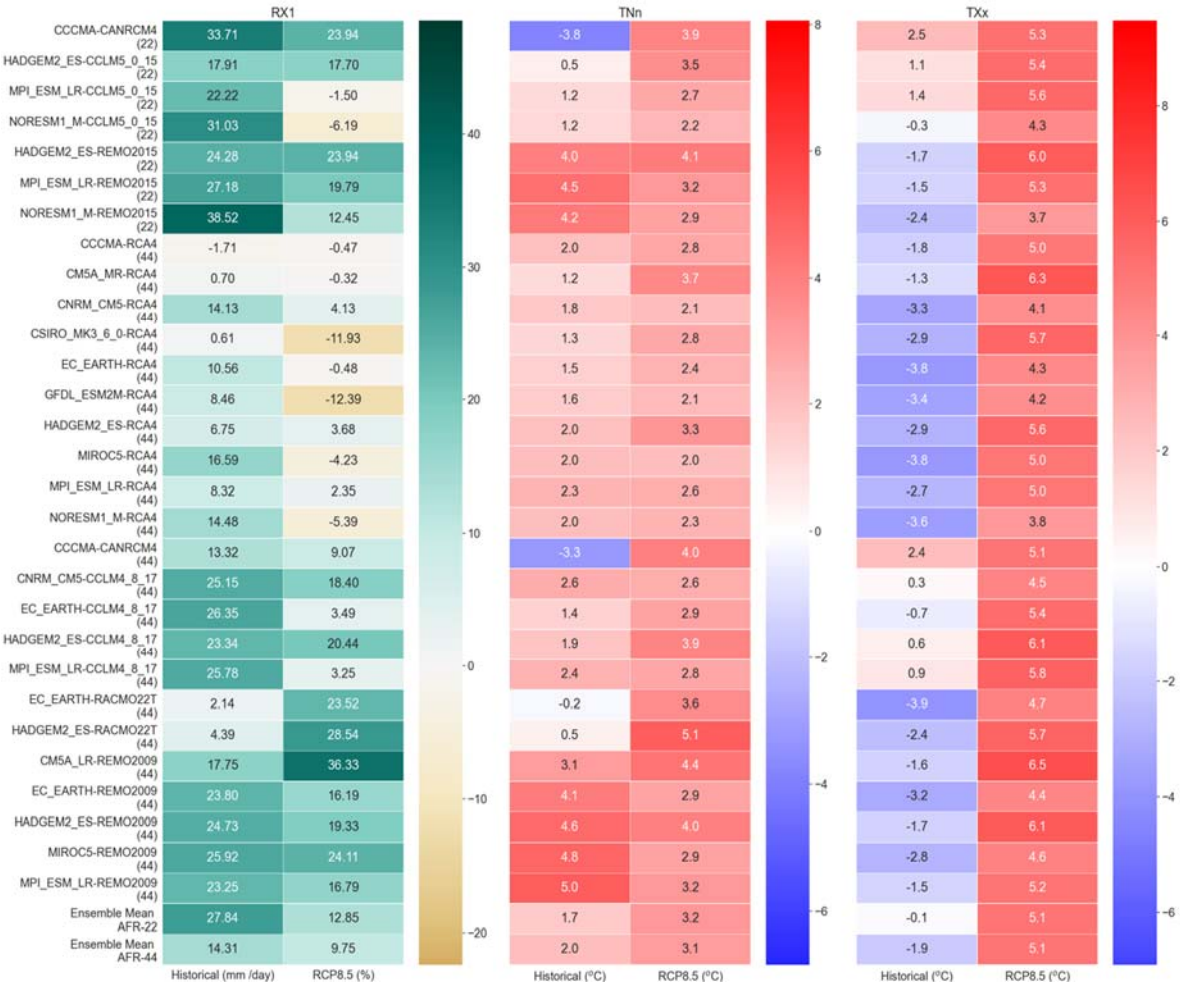


Figure 60: Averages for extreme variables (TXx, TNn and RX1) over South Africa. With Historical representing their bias compared to observations (CPC for TXx and TNn, and CHIRPS for RX1) and RCP8.5 comparing their end of century climate projection (2070 – 2100) to their historical climatology (1975 – 2005).

Zambia

TXx over Zambia is well represented by models at all resolutions. The higher resolution model does have a warmer bias, with CCLM5 downscaling being furthest from observations (2 – 3 °C) and an ensemble average of 1.9 °C warmer from observations. The CORDEX AFR-44 are variable with RCA4 and RACMO22T simulations underestimating TXx and the remaining simulations from the other RCMs slightly overestimating this index. The ensemble average agrees with the observations. TNn, although not indicated by the ensemble average, is also well represented by most models. Here, the REMO simulations at both resolutions, underestimate TNn significantly (+ 5.6 °C to 7.1 °C). When removing these simulations from the ensemble average, CORDEX AFR-22 and CORDEX AFR-44 do perform better (0.83 °C and 2.1 °C respectively).

RX1 is overestimated by over 35 mm day^{-1} in all the CORDEX AFR-22 models, with CCCMA-CANRCM4 overestimating by 67.82 mm day^{-1}. In the CORDEX AFR-44 ensemble, the CCLM4 and REMO2009 models both overestimate RX1, much like their higher resolution versions. RCA4 is closest to observations (+ ~0 – 10 mm day^{-1}), and the RACMO22T simulations underestimates RX1 by 14 mm day^{-1}.

Projections show an increase of between 2.5 °C and 5.9 °C for TNn, 3.6°C – 7°C for TXx and 1 – 47.9% for RX1 (Figure 61). There are some models in the high-resolution dataset that project a decrease in RX1, however, these are negligible (0.8 – 1.4 %).

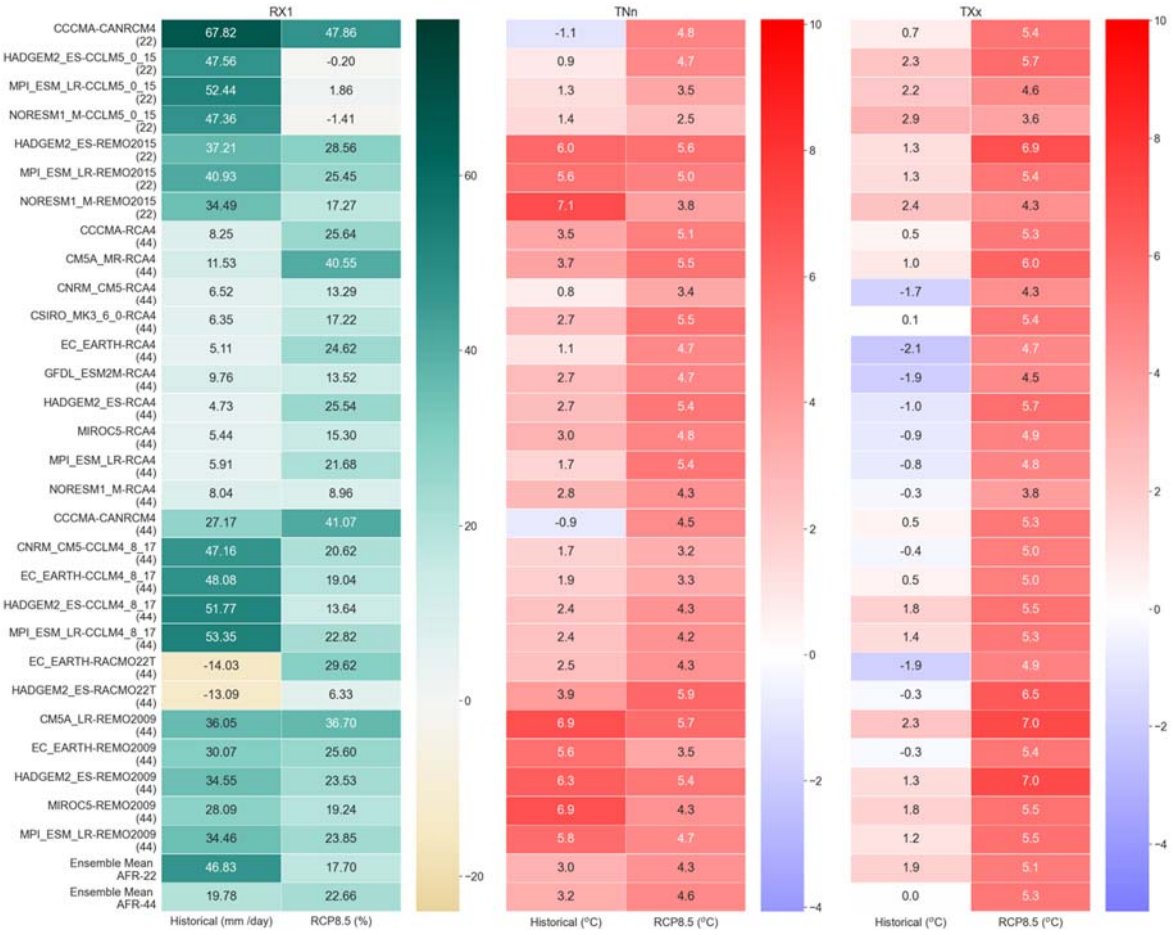


Figure 61: Averages for extreme variables (TXx, TNn and RX1) over Zambia. With Historical representing their bias compared to observations (CPC for TXx and TNn, and CHIRPS for RX1) and RCP8.5 comparing their end of century climate projection (2070 – 2100) to their historical climatology (1975 – 2005).

Mozambique

TXx is slightly overestimated by the CORDEX AFR-22 ensemble (+ 2.0 °C ensemble average), with CCLM5 downscaled simulations having the warmest bias (>2 °C). The CORDEX AFR-44 ensemble performs better, with an average just 0.3 °C higher than observations. RCA4 and RACMO22T are the only models to underestimate this variable. TNn is again overestimated by the CANRCM4 simulations at both resolutions (> -2°C), while REMO runs all underestimate by 3 °C to 4.5 °C.

This country has one of the poorer representations of RX1. RACMO22T again has a dry bias and understates this value by 14 – 19 mm day^{-1}. REMO is again the wettest of the downscaling's across both resolutions with all simulations exceeding observations by >45 mm day^{-1}, with the worst representation being off by over 80 mm day^{-1}. The RX1 projections are varied with projected reductions of ~9% and increases of 40.78%. RCA4 simulations project the greatest decrease in RX1. The degree of increase is not as clear with projections ranging from a few mm per day to over 40 mm day^{-1}. All models agree that TXx will increase and TNn will be warmer in the future, with increases in both of between 2 °C and 6 °C (Figure 62).

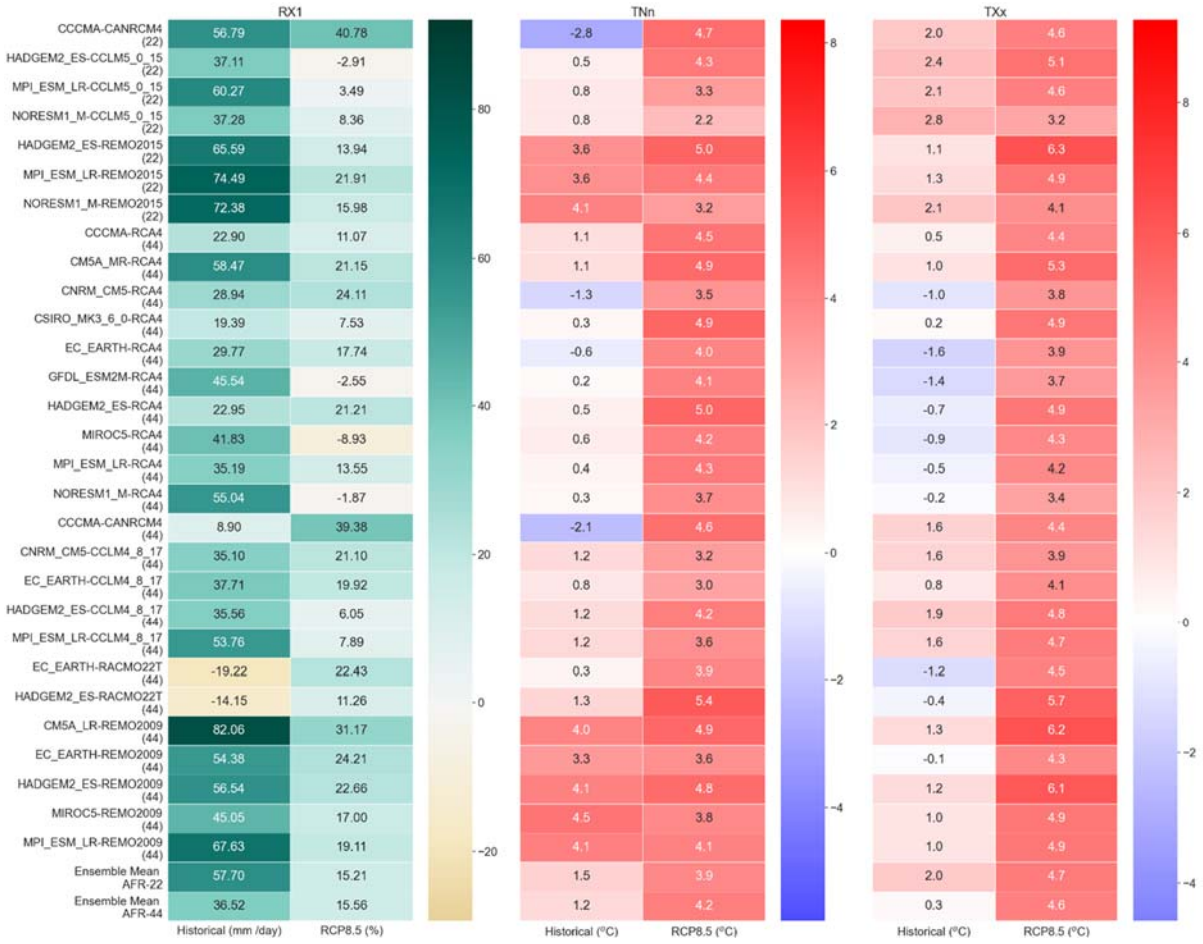


Figure 62: Averages for extreme variables (TXx, TNn and RX1) over Malawi. With Historical representing their bias compared to observations (CPC for TXx and TNn, and CHIRPS for RX1) and RCP8.5 comparing their end of century climate projection (2070 – 2100) to their historical climatology (1975 – 2005).

4.2.2.2.2 Limitations

Here the limitations of the CORDEX ensemble are discussed (Table 7) as well as the ones from CORDEX AFR-44 (Table 8). This provides a comprehensive overview of the limitations when reproducing and projecting extreme events (TXx, TNn and RX1).

Table 7: Limitations of the CORDEX AFR-44 RCM dataset,

Limitations	Impact
Weak dependence on GCM	The CORDEX ensemble (at both resolutions) appears to have a weak dependence on the driving GCM, with the RCM downscalings' showing very similar spatial and quantitative results in particular for the precipitation index, regardless of the driving GCM. This has also been noted in other studies (Dosio et al., 2019, 2021; Gnitou et al., 2021).
Reduction in spread / uncertainties	Due to the similarities in RCM iterations, the range of uncertainty is reduced.

Temperature representation	Although some models do perform well, there is still a cold bias for TXx and a warm bias for TNn, although generally TXx is better represented. This was also found in a study of CORDEX over other regions (Yu et al., 2021), where the representation of TXx was more skilful than TNn. It has also been shown that some models in this ensemble exhibit a temperature bias over the ITCZ due an inability to correctly represent clouds (Kothe, Panitz, & Ahrens, 2014).

Table 8: Limitations of the CORDEX AFR-22 RCM dataset

Limitations (CORDEX AFR-22)	Impact
Small ensemble size	The limited ensemble size of the CORDEX AFR-22 set limits its range of uncertainty. This leads to inevitable skewing of distributions of climate change responses should there be an outlier in the ensemble.
Much wetter bias than coarser resolution model	This ensemble was considerably wetter than CORDEX AFR-44 models (~20+ mm day^{-1}) over southern Africa. With precipitation data, the limited gauge data available cannot be ignored, resulting in observational uncertainty. With the data being of a higher resolution it is likely to pick up more of this uncertainty and this will affect the perceived bias, much like that seen in Gnitou et al (2021).
Limited added value compared to CORDEX AFR-44	Publications have already shown that the CORDEX AFR-44 dataset is able to represent extreme events well (Omar & Abiodun, 2017; Abiodun et al., 2020; Diallo et al., 2015; Pinto et al., 2016). Here CORDEX AFR-22 does not seem to show an improvement in comparison with CORDEX AFR-44 being outperformed by the coarser dataset at times. This could be due to several factors, including the limited size of the CORDEX AFR-22 dataset. However, there have also been studies that suggest an increase in horizontal resolution alone does not lead to an increase in performance, as these models need to be tuned and re-configured to work at this resolution to see an improved performance (Panitz et al., 2014; Sørland et al., 2021).

4.2.2.3 CCAM

The Conformal-Cubic Atmospheric Model (CCAM) is a variable-resolution global climate model developed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO, McGregor, 2005). CCAM is coupled to a dynamic land-surface model CABLE (CSIRO Atmosphere Biosphere Land Exchange model). The model was applied at the Global Change Institute in South Africa, as part of the FOCUS-Africa project, to downscale ERA-Interim reanalysis data to an 8 km spatial resolution over southern Africa, for the period 1979-2017. The experimental design starts by nudging CCAM in the ERA Interim data to obtain 50 km quasi-uniform resolution simulations (as described by Horowitz et al. (2017)) - forced at its lower boundary with the bias-corrected sea-surface temperatures (SSTs) and

sea-ice concentrations (SICs) from a CMIP5 GCM simulation. No atmospheric forcing from the GCM is applied. This approach avoids biases from the GCM, like for example the common Pacific cold tongue bias or the overestimation of SSTs along the west coast of southern Africa. The 8 km resolution runs over southern Africa were spectrally nudged within the output of the 50 km resolution global simulations following the approach of Engelbrecht et al. (2019).

A second set of CCAM simulations analysed has been obtained by downscaling the projections of future climate change of 6 CMIP5 GCMs under a low mitigation scenario, Representative Concentration Pathway 8.5 (RCP8.5), for the period 1961-2100. The GCM simulations were first downscaled to 50 km resolution globally Archer et al. (2018), with additional details in (Engelbrecht et al. (2015) and subsequently to 8 km resolution.

4.2.2.3.1 Performance over case study countries

This model has a cold bias over much of Southern Africa, with the exceptions of Botswana and Namibia, which are warmer than the observation (D3.1, 2021). Precipitation shows a wet bias which peaks in the eastern escarpment and a dry bias across the Cape Fold Mountains. These biases are smaller in comparison with the other simulations considered in this study, very likely an effect of the use of observed SST in the historical simulation (D3.1, 2021). The simulations are indicative of general rainfall decreases over southern Africa for the period 2089-2099 for RCP8.5 with respect to 1995-2005 (D3.1). An important exception is summer, for which most models project rainfall increases over the eastern escarpment regions. In several of these simulations, rainfall growth stretch into southern Mozambique, and in one of them the increases stretch westwards into the southern African interior. Substantial rainfall reductions in winter rainfall are projected, consistently across the simulations. Drastic warming is projected for all seasons, exceeding 5°C over extensive portions of the interior in all the projections (D3.1). The strength of these warming patterns seems to follow the climate sensitivity of their driving GCMs, a feature expected from downscaled simulations for which a good degree of consistency in the climate change patterns of the driving GCM is expected.

Extreme indices from CCAM have not been estimated. Surface daily variables in standard format from this dataset are not yet available, however they are expected to be available for the development of climate scenarios in the remaining part of this work package.

4.2.2.3.2 Limitations

Table 9: Limitations of the CCAM dataset

Limitations	Impact
Limited representation of GCM uncertainy	Although the 5 GCMs selected for the CCAM experiment give a good representation of GCM uncertainty on global scale, they do not necessarily represent the full uncertainty over Southern Africa for the indices considered in this study.
Diurnal cycle of precipitation not resolved	CCAM is still based on the use of parameterized convection, therefore it shares the same systematic issues as standard RCM and GCM (Nikulin et al, 2012)

4.2.3 Recommendations for subsequent tasks

Best practices in creating climate scenarios from multimodel datasets are mainly limited to GCM ensembles (e.g., seeMcSweeney et al., 2015). The inclusion of downscaling, with its additional uncertainty and its expected strong dependence on the driving GCM condition adds another dimension to the complexity of this problem. This is further increased when attempting to use *all* climate projections datasets in the development of climate scenarios. So here some recommendations, based on the evaluation done in this report and that of D3.1.

The first conclusion is that the CORDEX models at both resolutions should be used for climate projections over the case study countries. Here, the suggestion is to use as many of the ensemble members as possible to account for the widest range of uncertainties. When GCM-RCM combinations are duplicated across resolutions (e.g., CCCMA-CANRCM4 and the updated REMO and CCLM RCMs) it is plausible to only use the higher resolution members. Furthermore, for precipitation - due to the larger dependence on the RCM over the driving model - it is possible to take an ensemble average of all the GCM-driven simulations for each RCM, resulting in a single representation of each RCM. Some considerations that would need to be taken for this approach, would be the weighting of models such as RCA4, which makes up almost a third of the ensemble, as well as the reduction in range of uncertainty. This is an approach that could be tackled in many ways and add substantial, unnecessary uncertainty to the creation of climate scenarios. This problem, and the associated issue of how to integrate climate projections from different multimodel ensembles, is the subject of current, active, research, which will hopefully lead to the development of widely accepted best-practices for the use of these datasets in the future tasks.

4.3 Assessment of projections focusing on variables related to crop production

Climatic conditions play an important role in influencing the environment, society, economy, and the associated cultural practices. As the climate system of southern Africa countries becomes warmer (WMO, 2019; 2021), the region becomes also more vulnerable to changes in temperature and precipitation. In particular, crop production and food security are at great risk. The economies of southern African countries and their rural households depend largely on agriculture (ADB, 2019), making this sector a key component of growth and development. However, agriculture in the region is largely based on smallholder farming, with low use of agricultural inputs including irrigation, fertilizers, and pesticides. This makes crop production extremely sensitive to climate fluctuations, which may thus hamper livelihoods, food and water security and long-term development in the region (WMO, 2021).

In the future, crops are expected to face episodes of extreme temperatures and changes in the timing and characteristics of the rainy season, including frequency and distribution of rainy days (IPCC, 2019; 2021). Models project a general increase in precipitation intensity and decrease in frequency, resulting in more dry days and longer dry spells (Funk et al. 2019; Giorgi et al. 2019; Allan et al. 2020; Wainwright et al. 2021). Since reductions in water availability limit crop growth, longer dry spells coupled with the

shift in the timing of the rainy season may have negative impacts on crop yields and food production (Rockström et al. 2010).

Typically, climate observations correspond to quantitative variables with a direct comprehensible physical meaning such as total precipitation received over a defined period, daily maximum, or minimum temperature. Yet the environmental, societal, and economic impacts of climate features in crop production, depend on specific climate conditions or event characteristics over a time-period. These conditions or events include the timing of the rainy season, the long-term cumulative climate conditions, the extreme values recorded over a period, the frequency of days with specific characteristics and the frequency of events above or below specific thresholds (Wainwright et al 2017; Hariadi et al, 2022; Nkrumah et al, 2022). To capture and quantify the specificity of these climatic impact factors, much effort has been put into developing agroclimatic indicators which can be employed on different natural systems and socioeconomic sectors (Giorgi et al. 2019; Wainwright et al. 2021). Particularly relevant to crop production systems are heat stress under night and day time, late onset and shorter duration of the rainy season, and length of dry spells (Harrison et al. 2011; Gitz et al., 2016; Sun, et al. 2019; Shah et al. 2021).

General Circulation Models (GCMs) from the Coupled Model Intercomparison Project (CMIP) are the most credible tools to project future climate (Randall et al., 2007). Studies based on CMIP models show that the representation of agroclimatic indicators are in many ways model dependent (James et al, 2008; Moise et al, 2015). Consequently, it is very relevant to organize studies aimed at evaluating the performance of CMIP6 models in regard to the accurate representation of agroclimatic indicators that are relevant for crop production. The ability to simulate climate of individual CMIP6 models can vary depending on which aspect of a model simulation is considered (Taylor et al. 2012; Eyring et al., 2016). Although it may be difficult to select the best performing models, it is possible to identify a subset of models that have a higher accuracy in representing critical aspects of the climate such as parameters relevant for crop production (Evans et al 2013; CSIRO and Bureau of Meteorology, 2015).

In this section, the ability of eight CMIP6 historical climate models in accurately representing agroclimatic indicators relevant for crop production in the southern Africa region is examined. The growing season agroclimatic indicators include number of rainy days, number of light-to-medium and very heavy rainy days, longest dry spell, number of extreme hot days and nights and the rainy season calendar parameters (onset, cessation, and duration). These indicators are critical to managing crop production processes and directly relevant to decision support for strategic planning of crop production for years to come. Such information on performance of individual models is relevant when users are choosing a subset of models for application in impact assessment and to define the confidence in projections. In particular, the findings of similar studies are relevant for crop producers, breeders, and government policy makers. At the same time, the results from the model evaluation are very important when choosing host models for dynamical downscaling.

4.3.1 Study Area

This study focuses on evaluation of CMIP6 model performance over the southern Africa domain. This area is predominantly a unimodal rainfall region in which the rainy season spans from October to April.

The target unimodal rainfall region was determined by objective empirical orthogonal function (EOF) analysis of continuous monthly total rainfall (Figure 63).

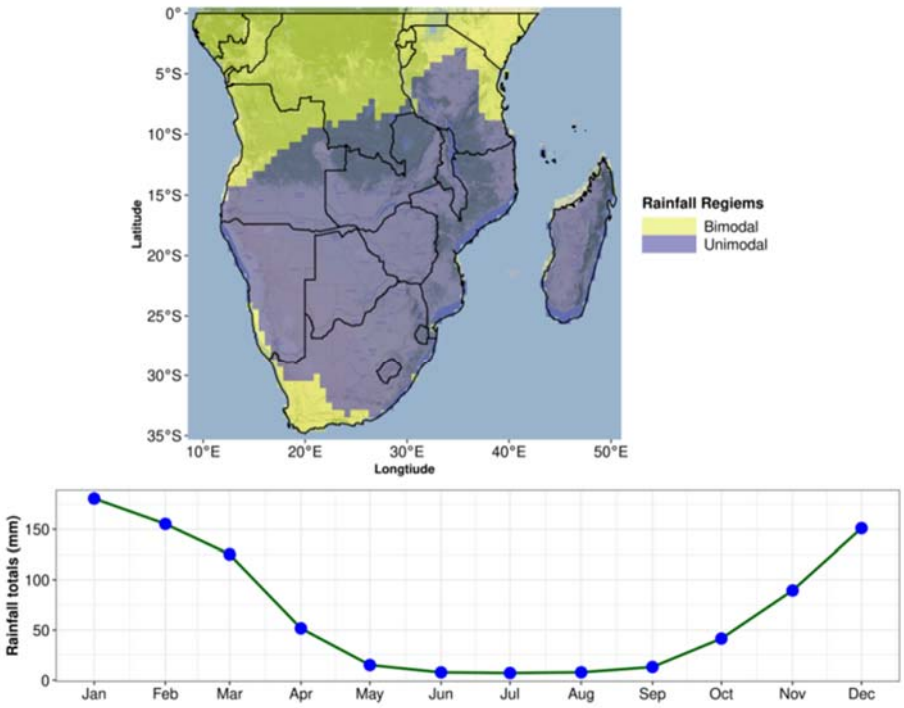


Figure 63: The target study area, presented as a unimodal rainfall regime *(top panel)*; the annual cycle of mean monthly rainfall totals for the unimodal regime (*bottom panel).*

4.3.2 Data and methods

4.3.2.1 CMIP6 Models

The Coupled Model Intercomparison Project (CMIP6) daily historical experiment data (Eyring et al., 2016), which cover the period from 1980 to 2010, has been obtained from the Climate Data Store (CDS) of the Copernicus Climate Change Service (C3S). The GCMs that have been evaluated in this study are shown in Table 10 together with their horizontal resolutions and modelling centers. The GCM outputs employed in this study include daily maximum near-surface air temperature, daily minimum near-surface air temperature and daily precipitation. Models have been selected based on the availability of the required daily data for historical period.

Table 10: Description of the CMIP6 climate models used in the study.

Model Name	Modelling Centre	Horizontal Resolution	Reference
BCC-ESM1	Beijing Climate Center (BCC)	The model was run in native nominal resolutions: (i) atmosphere, 250 km (ii) atmospheric chemistry, 250 km (iii) land, 250 km (iv) ocean, 50 km and (v) sea ice, 50 km.	Tongwen Wu et al., 2020
CanESM5	CCCMA (Canadian Centre for Climate Modelling and Analysis)	The model was run in native nominal resolutions: (i) aerosol, 500 km (ii) atmosphere, 500 km (iii) atmospheric chemistry, 500 km (iv) land, 500 km (v) land ice, 500 km (vi) ocean, 100 km (vii) ocean biogeochemistry, 100 km and (viii) sea ice, 100 km.	Swart et al., 2019
CMCC-ESM2	CMCC (Centro Euro-Mediterraneo per I Cambiamenti Climatici)	The model was run in native nominal resolutions: (i) aerosol, 100 km (ii) atmosphere, 100 km (iii) land, 100 km (iv) ocean, 100 km and (v) sea ice, 100 km.	Lovato et al., 2022
CNRM-CM6-1-HR	CNRM-CERFACS (National Center for Meteorological Research, Météo-France and CNRS laboratory)	The model was run in native nominal resolutions: (i) aerosol, 100 km (ii) atmosphere, 100 km (iii) atmospheric chemistry, 100 km (iv) land, 100 km (v) ocean, 25 km (vi) sea ice, 25 km.	Voldoire et al., 2019
EC-Earth3-CC	EC-Earth-Consortium	The model was run in native nominal resolutions: (i) atmosphere, 100 km (ii) land, 100 km (iii) ocean, 100 km and (iv) sea ice, 100 km.	(Ralf Döscher et al., 2022)
EC-Earth3-Veg-LR	EC-Earth-Consortium	The model was run in native nominal resolutions: (i) atmosphere, 250 km (ii) land, 250 km (iii) ocean, 100 km and (iv) sea ice, 100 km.	(Ralf Döscher et al., 2022)

GFDL-ESM4	NOAA-GFDL (National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory)	The model was run in native nominal resolutions: (i) aerosol, 100 km (ii) atmosphere, 100 km (iii) atmospheric chemistry, 100 km (iv) land, 100 km (v) land ice, 100 km (vi) ocean, 50 km (vii) ocean biogeochemistry, 50 km and (viii) sea ice: 50 km.	(Dunne et al., 2020)
MIROC6	MIROC (Atmosphere and Ocean Research Institute (AORI), Centre for Climate System Research - National Institute for Environmental Studies (CCSR-NIES)	The model was run in native nominal resolutions: (i) aerosol, 250 km (ii) atmosphere, 250 km (iii) land, 250 km (iv) ocean, 100 km and (v) sea ice: 100 km.	(Hiroaki Tatebe et al., 2019)

4.3.2.2 Evaluation Dataset

At the core of every model evaluation there is a reference dataset to which model simulations can be compared. In this study, the daily gridded climate dataset from AgERA5 (obtained from the European Copernicus Program, AgERA5, 2021) is employed to evaluate GCMs considering the maximum air temperature, minimum air temperature, rainfall, and solar radiation.

4.3.2.3 Agroclimatic indicators

In this work the focus is on the agroclimate indicators that could bring severe loss to crop yields, such as growing region exposure to critical thresholds and events above which yield declines occur (Zhu and Troy, 2018). These indicators have been widely used to study observed and modeled climate variability and impact on crop production across the globe (Wilson et al. 2022; Kim et al. 2020; Wainwright et al. 2022). They can be found in Table 11.

Table 11: Description of agroclimatic indicators used in the study

Indicators	Description	Definition	Units
NRD	Number of rainy days	Days when rainfall is >= 1mm	Days
Rle20	Number of light to medium rainfall days.	Days when rainfall is at most 20mm	Days

Rge20	Number of very heavy rain days	Days when rainfall is at least 20mm	Days
LDS	Longest dry spell	Maximum number of consecutive dry days with PR < 1 mm	Days
TXge35	Number of extreme hot days	Number of days when Tx >= 35 °C	Days
TNge20	Number of extreme hot nights	Number of days when Tn >= 20 °C	Days
onsetDOY	Onset day of the rainy season	Onset day of the rainy season in day of the year (DOY)	DOY
cessationDOY	Cessation day of the rainy season	Cessation day of the rainy season in day of the year (DOY)	DOY
seasDur	Duration of the rainy season	Duration of the rainy season	Days

4.3.2.4 Definition of the onset and cessation of the rainy season

Several climatic factors such as seasonal rainfall amount, intraseasonal rainfall distribution and dates of onset and cessation of the rains influence crop yields and determine the crop production calendar. In particular, the onset of the rainy season appears to be the most crucial information for crop production management since it determines the planting period (Marteau et al., 2011). Indeed, a slow rainy season onset is a reliable harbinger of drought in most food insecure regions of Sub-Saharan Africa (Shukla et al. 2021). Numerous definitions exist for identifying the onset and cessation of the rainy season. In past studies, the onset of the rainy season has been either considered as the start of the rainy season, and hence was identified based on rainfall data, or was identified based on an agroclimatic approach. The latter strategy defines the onset of the rainy season as the optimal date that ensures sufficient soil moisture during planting and early growing periods to avoid crop failure after sowing (Mugalavai et al., 2008; Marteau et al., 2011), and requires information both relative to rainfall and temperature as well as to soil features of the area under study.

A normal cropping period is defined as one when there is an excess of precipitation over potential evapotranspiration (PET). Such a period meets the evapotranspiration demands of crops and recharges the moisture of the soil profile (FAO 1978; 1977; 1986). Thus, onset will start on the first day after 01-Sep when the actual-to-potential evapotranspiration ratio (Ea/Ep) is greater than 0.25

and is followed by a 20-day period in which plant available water (PAW) remains above 10mm. The logic behind is a consistent available soil moisture of 10 mm and above for the plant to extract will support vegetative growth (Mugalavai et al., 2008). Similarly, the rainy season will end on the first day after 01-Mar when Ea/Ep <= 0.25 and followed 12 consecutive non-growing days (PAW < 10mm).

4.3.2.5 Evaluation metrics

To allow comparing all data products, both CMIP6 and AgERA5 data have been resampled to a common 0.5° × 0.5° grid using a bi-linear remapping procedure using a 'raster' package in R statistical programming environment. The analysis in this study is carried out with respect to the historical period 1980 - 2010. The performance metrics represents the ability of the CMIP6 models to accurately capture and reproduce the observed spatial pattern (seasonal climatology) and temporal variability (inter-annual) of agroclimatic indicators over the study area.

Percent bias (PBIAS) has been used to evaluate the performance of the CMIP6 models in simulating the spatial pattern of seasonal climatology of agroclimatic indicators. The PBIAS measures the average tendency of the model values to be larger or smaller than the observed ones. The optimal value of PBIAS is 0.0, with values closer to zero indicating better model performance. Positive values indicate overestimation bias, whereas negative values indicate model underestimation bias.

$$PBIAS = 100 * \frac{\sum_{i=1}^{n}(\hat{x}_i - x_i)}{\sum_{i=1}^{n} x_i}$$

Where $\hat{x}_i$ is the seasonal summary, totals over November to April period, of agroclimatic indicators from CMIP6 models and x_i is indicators from observations for the same period.

Taylor diagrams have been employed to quantify the degree of correspondence between CMIP6 model simulations and observation, taking into consideration the inter-annual variability of seasonal summary of indicators, which are shown by correlation coefficients, root mean square error (RMSE), and the normalized standard deviations (Taylor 2001). When the correlation coefficient and the standard deviation are close to 1 and the RMSE is close to 0, this is indicative of the best match between observation and model simulation.

To evaluate the inter-annual variability of the country averaged seasonal summary of agroclimatic indicators, the root mean square error (RMSE), and the index of agreement have been computed. The index of agreement (d) is a standardized measure of the degree of model prediction error and varies between 0 and 1 (Willmott, 1981). A value of one indicates a perfect match, while 0 indicates no agreement. The index of agreement can detect additive and proportional differences in the observed and simulated means and variances. However, it is overly sensitive to extreme values due to the squared differences (Moriasi et al., 2007).

$$d = 1 - \left[\frac{\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}{\sum_{i=1}^{n}\left[\left(\left|\hat{x}_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\right| + \left|x_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\right|\right)^2\right]} \right]$$

Where $\hat{x}_i$ is the seasonal summary, totals over November to April period, of agroclimatic indicators from CMIP6 models and x_i is indicators from observations for the same period, whereas 'i' runs from 1981 to 2010 (n = 30 years).

4.3.3 Results

4.3.3.1 Spatial Pattern

The climatic patterns of the indicators used are shown in Figure 64 andFigure **65**, which provide a measure of the seasonal behavior of the area under analysis. The number of rainy days (NRD), including light to medium (Rle20) and very heavy rainy days (Rge20) exhibits a decreasing gradient when going from the north (humid) and northeastern (sub-humid) part to southwestern areas (arid) with the lowest values occurring over the shore west of the Kalahari Desert, and the highest values occurring over the Congo rainforest. Dry spells of length greater than 8 days are common on the arid and semi-arid parts of central Tanzania, southern Mozambique, parts of Zimbabwe and Botswana, eastern South Africa, and most of Namibia. Those areas also show a frequent number of extreme hot days (Figure 64). Such prolonged dry spells coupled with frequent extreme hot days and nights may lead to significant stress on crop growth resulting in low crop productivity and hampering food security in the area.

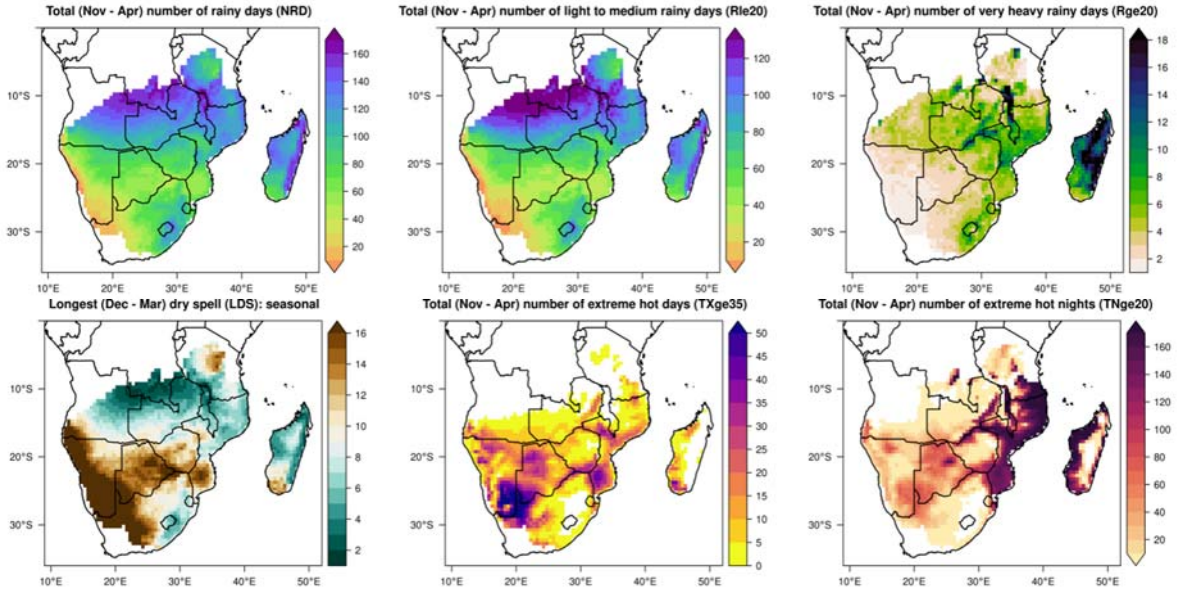


Figure 64: Spatial pattern of mean seasonal agroclimatic indicators: climatology over 1981 to 2010 period.

The spatial pattern of the earliest, mean and late onset and cessation of the rainy season is also presented, as well as its shortest, average and longest duration during 1980 to 2010 period (Figure 65). To correctly inform cropping practices it is useful to know the typical or time-mean onset date for a particular location. Knowledge of a mean date allows the calculation of onset date anomalies. The mean onset of the rainy season occurs in the beginning of October over the southern part of DRC, Northern Angola and the southeastern part of South Africa. During November and December, it gradually moves towards the south, initiating in later seasons for the countries at higher southern

latitudes. By late January and February, mean onset is achieved towards all the southern and southeastern arid and semi-arid areas. On average, the rainy season ends starting from late April at the eastern and southwestern parts, and the cessation progresses over time to the northwestern and western regions, until late July.

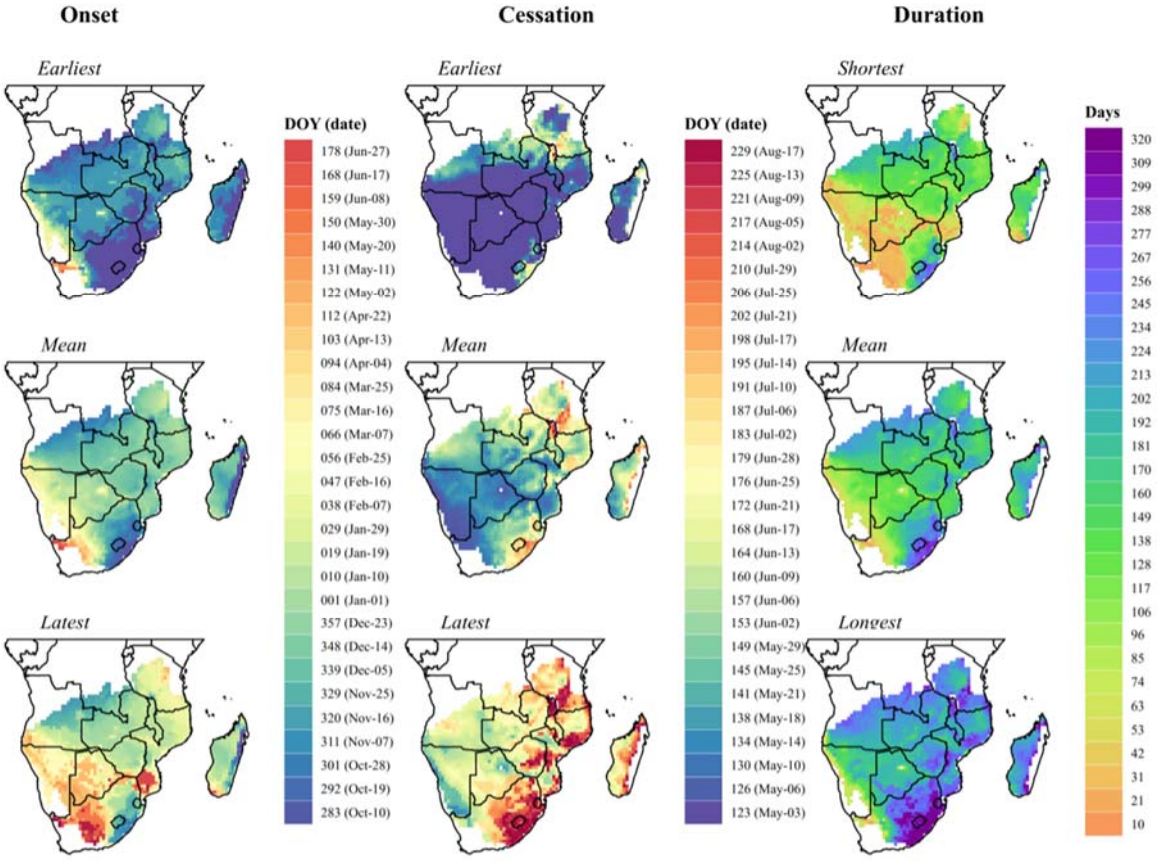


Figure 65: Spatial pattern of the earliest/shortest, mean, latest/longest onset cessation and length of growing season: climatology over 1981 to 2010 period.

The simulated seasonal climate means of 9 agroclimatic indicators have been compared with observation for the reference period of 1980 - 2010. The percentage bias of seasonal agroclimatic indicators by CMIP6 models is given in Figure 66Figure 67.. Overall, most models overestimate the number of rainy days (NRD), number of light to medium rain days (Rle20) and underestimate the duration of the longest dry spell (LDS), except for CMCC-ESM2. The seasonal number of extreme hot days (TXge35) is overestimated by BCC-ESM1, GFDL-ESM4 and MIROC6, while the rest of the models underestimate it. Both those models and CanESM5 overestimate the mean seasonal total number of extreme hot nights (TNge20).

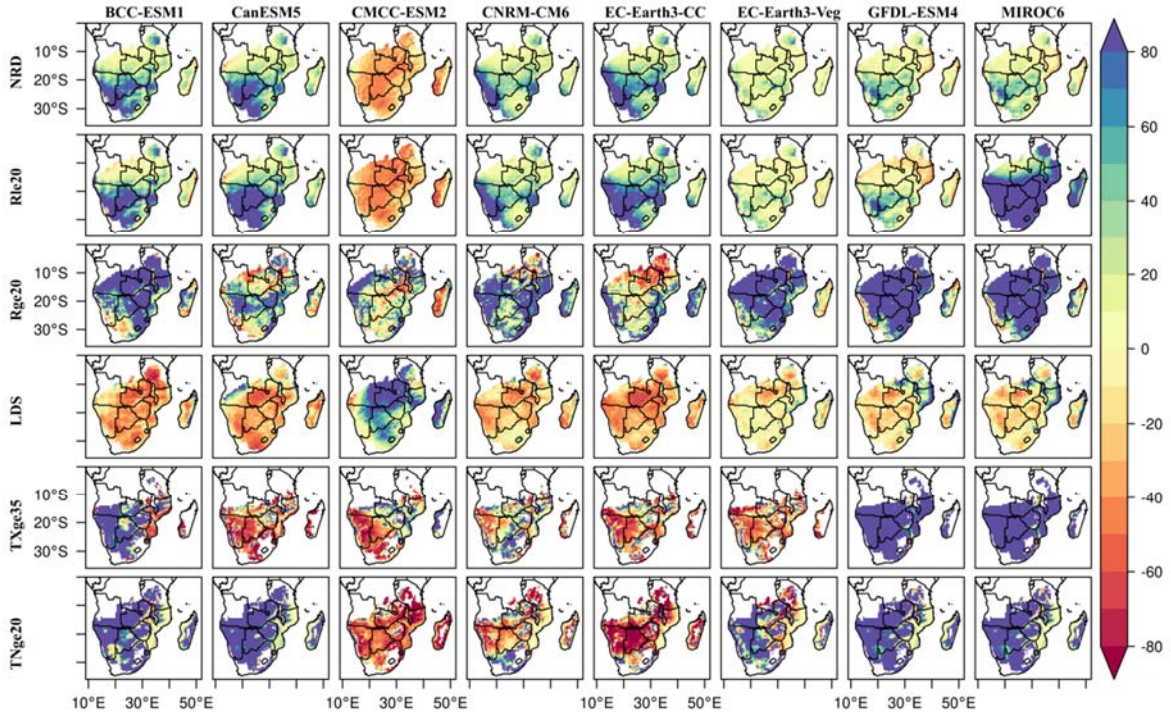


Figure 66: Percent bias (%) of CMIP6 agroclimatic indicators.

Most of CMIP6 models show consistency in estimating the mean onset, cessation, and duration of the rainy season. Generally, an early mean onset and longer mean duration of the rainy season is produced by most of the models except for CNRM-CM6 (Figure 67). The percentage bias for mean cessation of the rainy season is less variable across models than that of the onset.

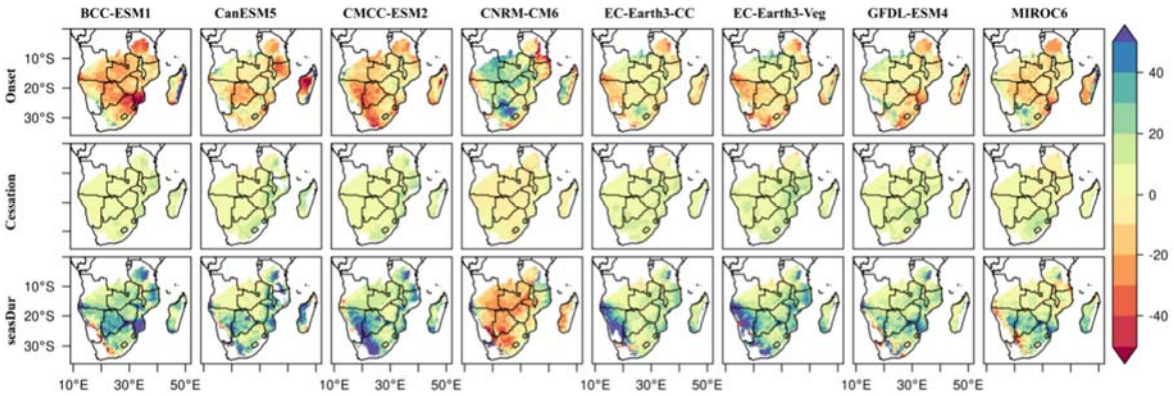


Figure 67: Percent bias (%) of CMIP6 agroclimatic indicators.

The performance of the CMIP6 models in simulating the spatial pattern of observed seasonal climatology is presented using a Taylor diagram (Figure 68).The diagram consists of three parts: i) interannual correlations indicated by black lines with labels on the outward extension of the line, ii) centered root-mean-square-error differences (CRMSE), indicated by labeled dark golden dashed contour lines and, iii) the ratio of variances (normalized standard deviation). The statistic is normalized by dividing both the CRMSE and the standard deviation of the model values by the standard deviation

of observations. The baseline reference point is indicated by a purple dot where correlation is 1 and CRMSE is 0. The more the model points are close to this purple point, the more they are similar to the reference.

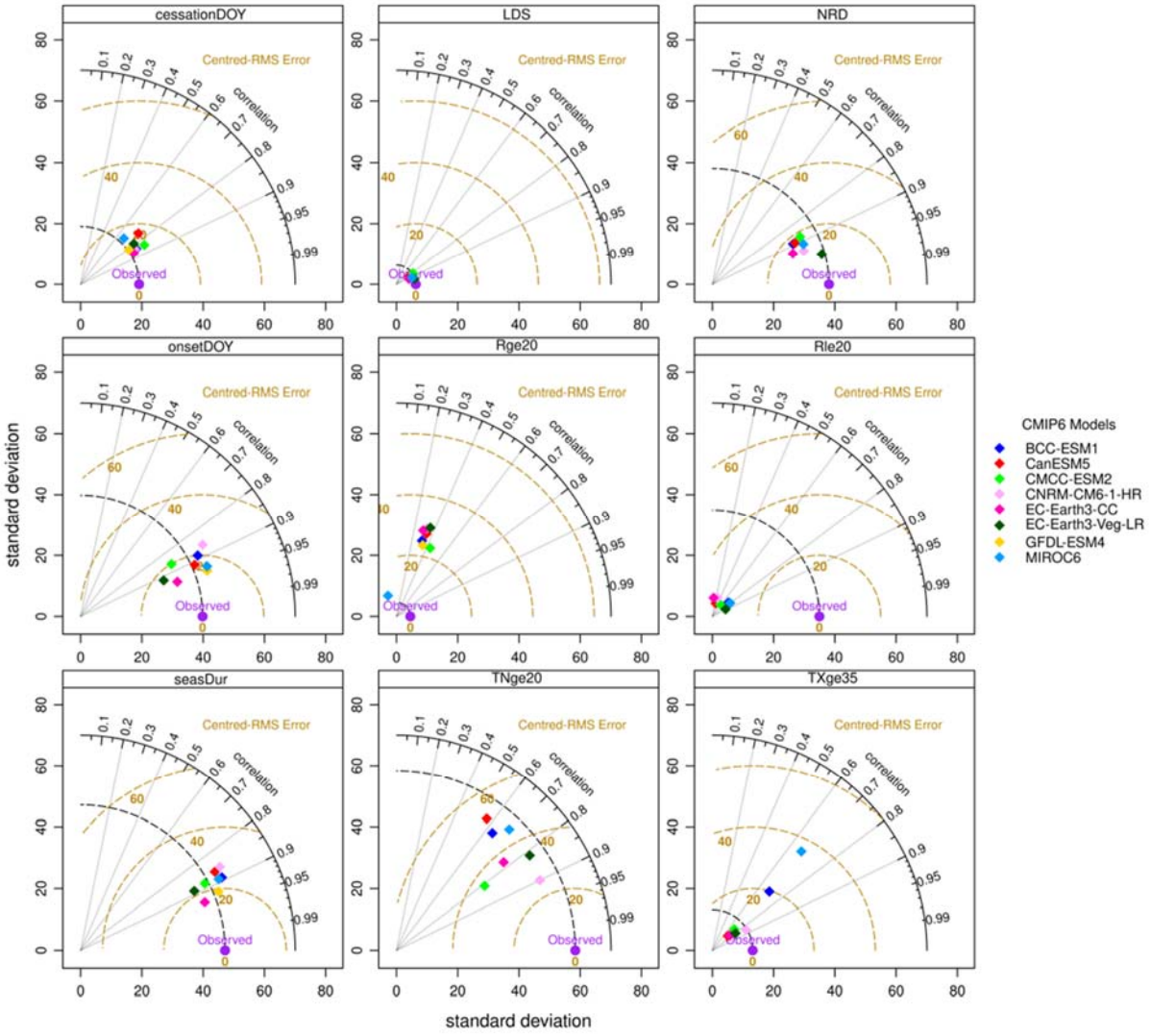


Figure 68: Taylor diagrams showing the performance of CMIP6 models in simulating the spatial pattern of observed seasonal climatology of agroclimatic indicators.

In general, all models show good agreement with observed values ($r > 0.6$) in representing the spatial pattern of the seasonal climatology. The spatial standard deviation of agroclimatic indicators simulated by the models is relatively higher, except for the simulation for the longest dry spell (LDS). Higher CRMS errors are associated with BCC-ESM1 and MIROC6 when simulating the number of extreme hot days (Txge35). It is also worth mentioning that the models struggled in simulating the number of extreme hot nights (TNge20).

4.3.3.2 Temporal variability

The performance of CMIP6 models in simulating the observed interannual variability of seasonal agroclimatic indicators can also be depicted with a Taylor diagram (Figure 69). Interannual correlations

are indicated by black lines, while interannual CRMSE are indicated by labeled dark golden dashed contour lines. The statistic is normalized by dividing both the CRMSE and the standard deviation of the model values by the standard deviation of the reference.

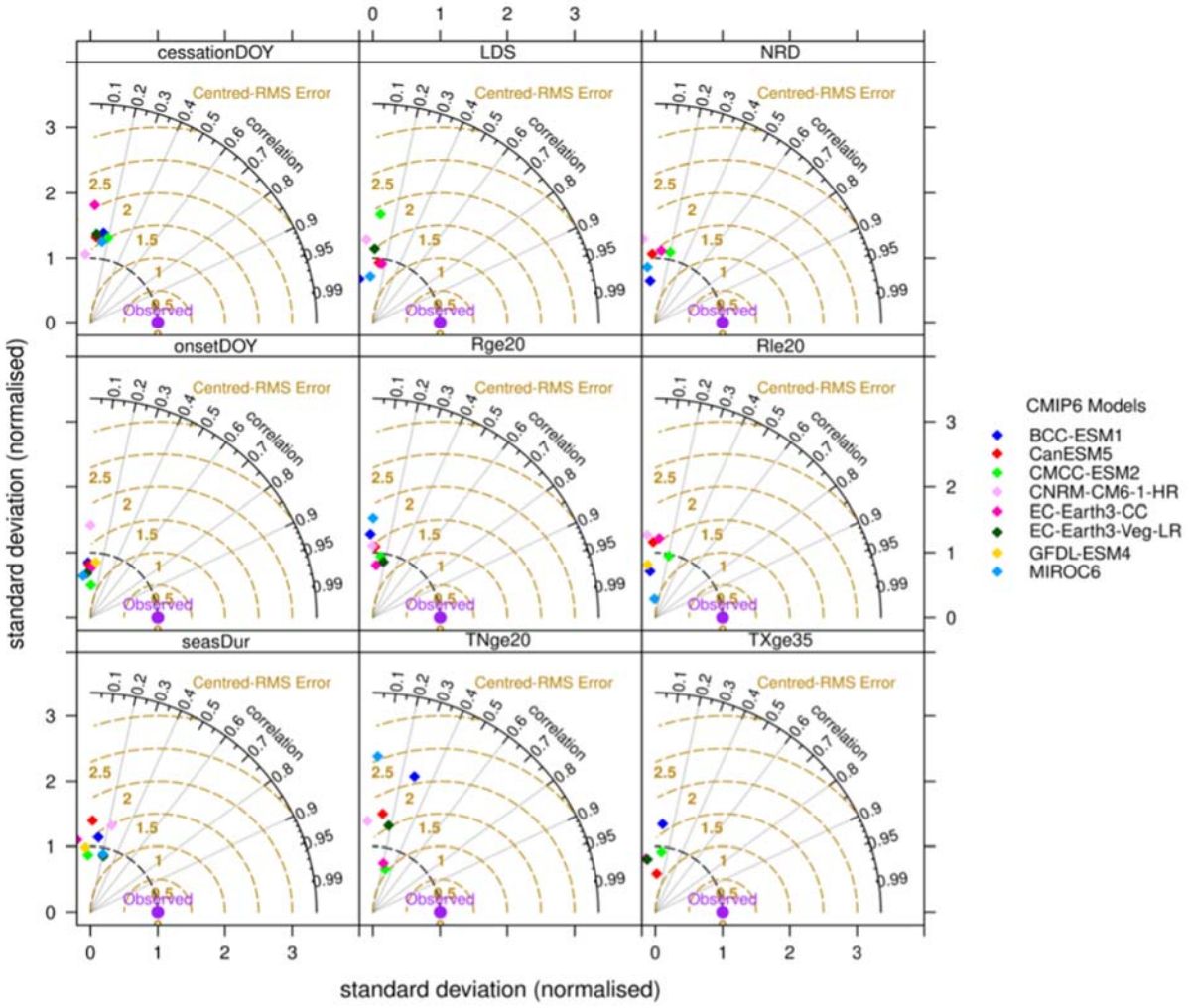


Figure 69: Taylor diagrams showing the performance of CMIP6 models in simulating the observed interannual variability (1981 - 2010) of seasonal agroclimate indicators.

The model values all have higher variation *(normalized standard deviation > 1)* than the reference and have lower levels of agreement *(r < 0.4)*. In general, the quality of the models is fair *(CRMSE < 2).* However, models such as BCC-ESM1 and MIROC6 show relatively higher error values in simulating *TNge20*, while EC-Earth3-CC have high errors in simulating *cessationDOY*. Relatively speaking, the models have a higher level of quality variation in simulating *TNge20* than for other agroclimatic indicators.

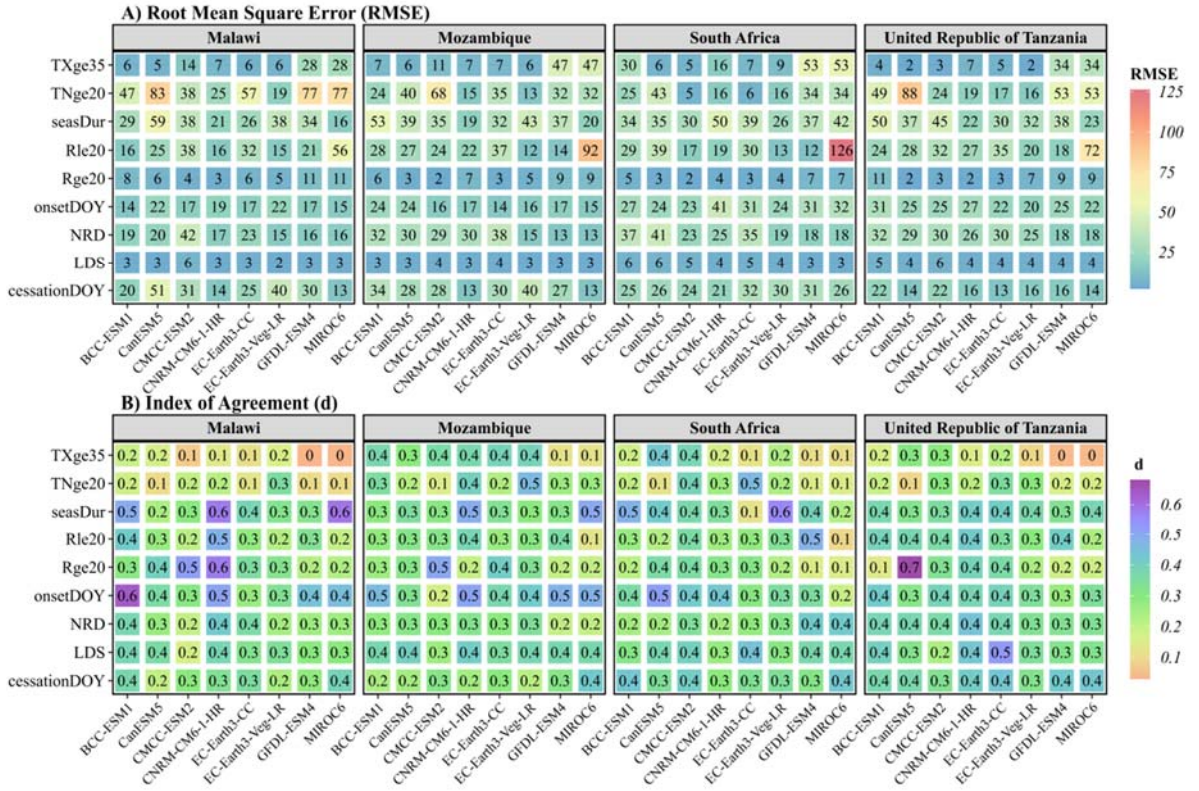


Figure 70: The 'portrait' diagram for evaluation score of country average agroclimatic indicators simulated by the CMIP6 models

The performance has been summarized for individual models for agroclimate indicators using a 'portrait' diagram (Figure 70). The portrait diagram depicts the magnitudes of RMSE and index of agreement for each country, showing average agroclimatic indicators by rows, and models by columns. The evaluation has been computed for year-to-year variation (1981 - 2010) of the growing season summary of agroclimatic indicators. The results show that most of the models perform considerably well in simulating most indices. Once again, high values of RMSE characterized the simulations for the number of extreme hot nights (TNge20) for all models except EC-Earth3-Veg-LR. Lower index of agreement in simulating seasonal number of extreme hot days (TXge35) has been observed over Malawi, South Africa and Tanzania.

4.4 Process based assessment of global and regional projections

The objective of this section is to identify projected changes in atmospheric processes that influence rainfall over Malawi. It is complementary to work done in Work Package 3 (WP3) that reports on moisture fluxes and rainfall having characterized the impact of climate variability and future change on hydropower generation in Lake Malawi and the Shire River basin.

Here the atmospheric processes have been considered at two scales, namely teleconnective and synoptic scales. These scales have been selected because teleconnections such as the El Nino Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) have been shown to affect rainfall in the Lake Malawi catchment region and synoptic scale processes facilitate the advection of moisture into the

region (as described in WP3, D3.2). The rainfall control of these processes has been quantified during the observed period and projected changes under the SSP245 emissions scenario.

4.4.1 Assessment of ENSO and IOD teleconnections in the Lake Malawi catchment

The aim of this section is to investigate whether CMIP6 models accurately capture teleconnections affecting rainfall in the Malawi catchment region, relative to that captured in the ERA5 reanalysis model. The Indian Ocean Dipole (IOD) and El Niño-Southern Oscillation (ENSO) indices have been quantified as they exert an interannual control on rainfall in the study region and evaluate the ability of CMIP6 GCMs to capture these controls and how they are projected to change in the medium term (2030 to 2059) under the SSP245 scenario.

4.4.1.1 Data and methods

The period 1961-2014 has been considered to ensure ERA5 and 24 CMIP6 historical runs are sampled over the same period, which provided a sample size of 669 3-month seasons over which to conduct the analysis of the events and an evaluation of the GCMs. The projected time horizon is chosen as 2030-2059 which constitutes the IPCCs medium time horizon; and SSP245 is selected as it represents the likely highest emissions scenario to be realized (as opposed to higher scenarios such as SSP585). Regarding the Lake Malawi catchment, it is defined as a box spanning 33E to 35.5E and 14.5S to 9.25S.

The Oceanic Nino Index (ONI) is used to define ENSO events. This is calculated as the 3-month rolling mean of Sea Surface Temperature area-weighted across the Nino 3.4 region [120W - 170W 5N - 5S]. While a warm (El Niño) event is defined when the SST anomaly, calculated against a rolling 30-year base period, exceeds 5 degrees for a minimum of 5 consecutive months, and a cold (La Niña) event when the SST is below 5 degrees.

The Indian Ocean Dipole (IOD) is defined as the area-weighted monthly SST difference between the tropical western Indian Ocean [50E - 70E, 10S - 10N] and the southeastern tropical Indian Ocean [90E - 110E and 10S - Equator]. Like the ONI, positive events are defined when the 3-month rolling mean of this index exceeds 0.5 degrees for a minimum of 5 consecutive months, and a negative IOD event is defined when the index is less than 0.5 degrees.

To study the impact these modes of variability have on rainfall in the Lake Malawi catchment, the identified periods where an event (+ve/-ve ENSO and +ve/-ve IOD), defined as occurring across 3 consecutive months, have been compared against the same 3 months when that event is not evident.

4.4.1.2 Results

The Lake Malawi catchment appears to be at the confluence of different anomaly modes in both IOD and ENSO (Figure 71). It is only during negative IOD and negative ENSO events that the catchment experiences statistically significant drier anomalies. Spatially, ENSO has a greater (more statistically significant) influence over the southern region of the catchment and the IOD, in the northern areas.

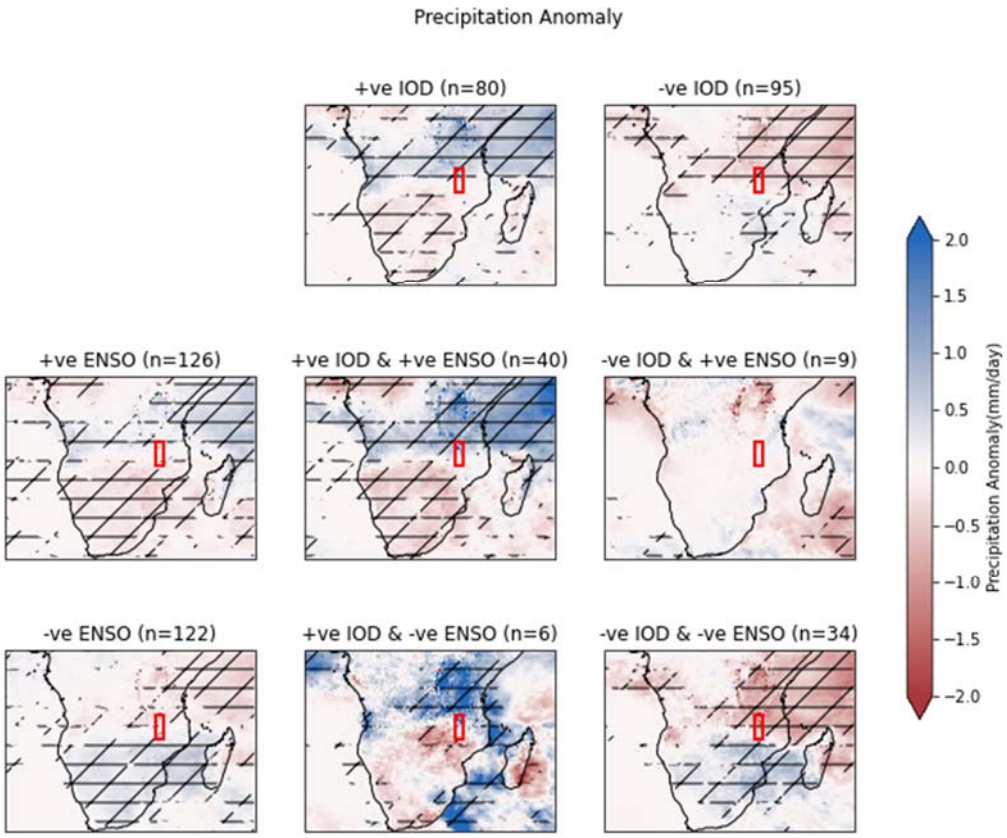


Figure 71: Precipitation anomalies associated with each event over the period 1961-2014. Cross hatching indicates significance at p<0.01. Red box indicates Malawi catchment.

The 24 GCMs are evaluated against the ERA5 for their ability to capture rainfall anomalies associated with each event (Figure 72). Spatial rainfall anomaly patterns associated with each of the event modes are generally captured in the GCM ensemble mean except for +IOD/-ENSO which produces a positive rainfall anomaly over the southern Africa interior, whereas in the ERA5 this is a negative anomaly. The magnitude of the anomaly is generally smaller in the GCM ensemble mean. Over the catchment region the rainfall anomaly patterns are very similar. Therefore, the conclusion is that the GCMs adequately simulate the spatial distribution of rainfall anomalies in the study region.

Projected rainfall anomaly patterns of each event mode for the period 2030-2059 under the SSP245 scenario is shown in Figure 73. The spatial distribution anomalies for each mode are similar to those in the historical simulations (Figure 73a) and differences between the two periods are low under the pure IOD modes (Figure 73b). It is only under the +IOD/-ENSO mode, which in the ERA5 data indicate much wetter than average anomalies, that there are larger differences between the historical and future periods. This suggests that in future such events may be wetter than in the historical period over the region. However, it must be noted these events are very rare (see next section and Figure 73).

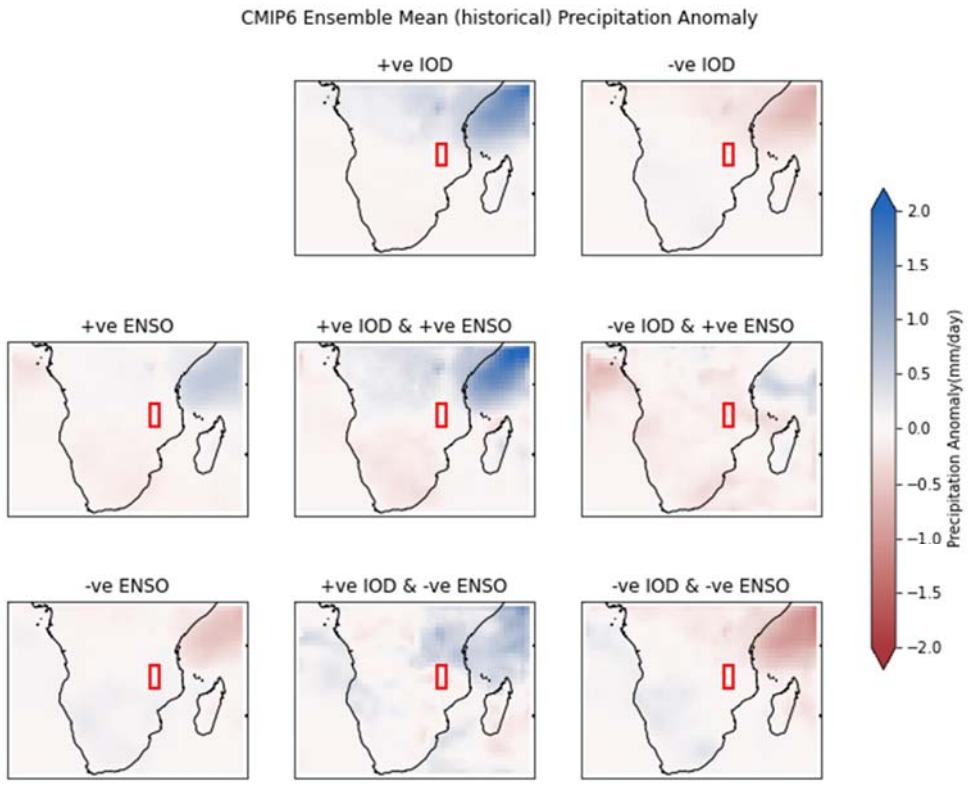


Figure 72: Precipitation anomaly of the 24-member GCM ensemble associated with each event over the historical period 1961-2014.

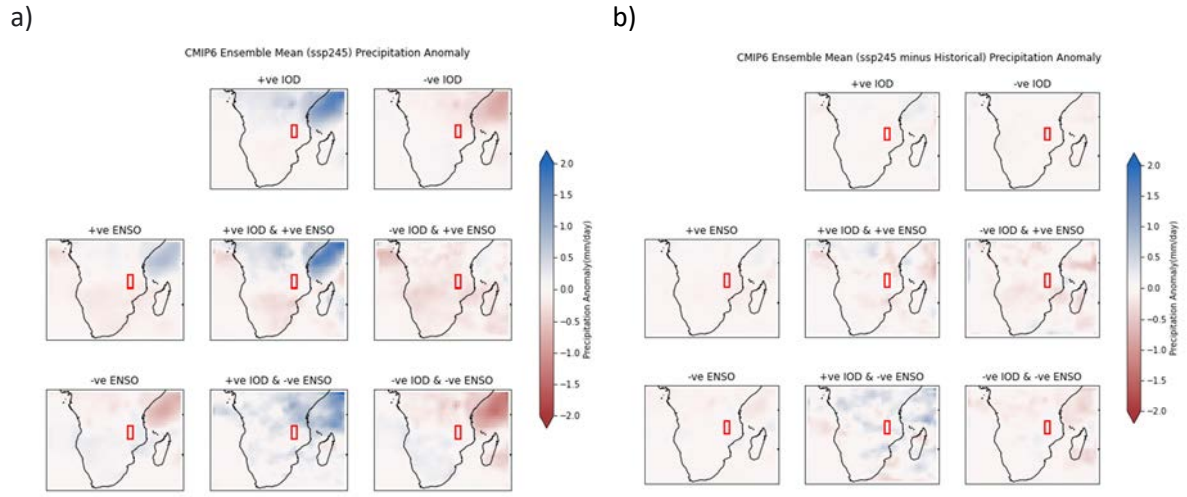


Figure 73: CMIP6 anomalies of rainfall during events for (a) the period 2030 to 2059 under the SSP245 scenario and (b) the difference between this and the historical period.

Although the spatial distribution of rainfall anomalies associated with the forcing modes are well represented by the 24 GCMs, it is still necessary to understand the local expression of rainfall under each forcing mode. This has been done through analyzing rainfall as a percentile of the full rainfall distribution for the region, associated with the occurrence and co-occurrence of events. The frequency

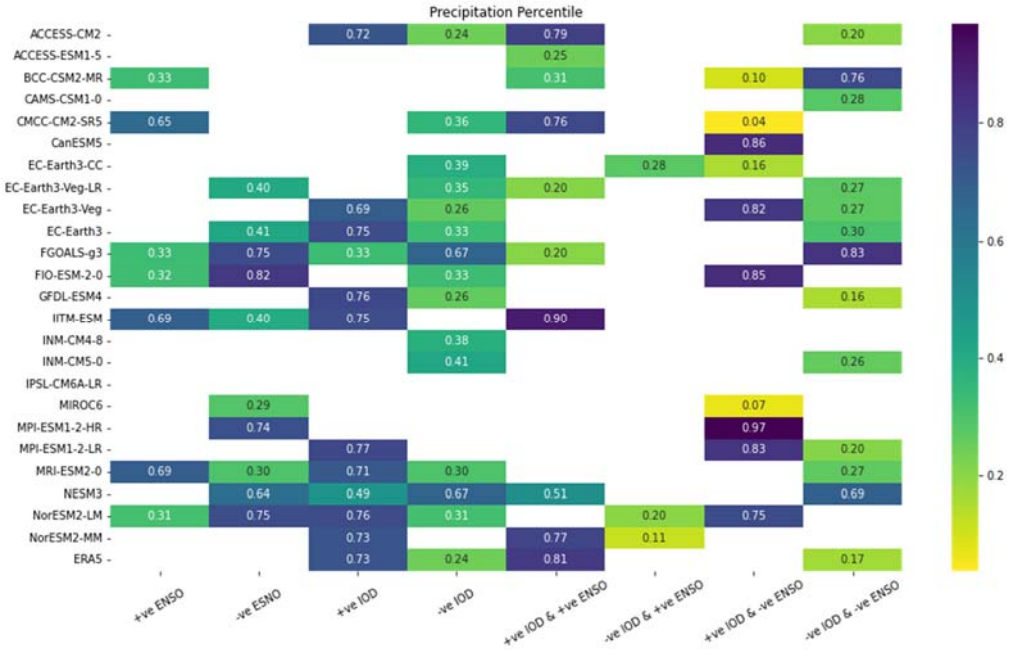
for each event in the model has also been identified and compared to the ERA5 frequency, and how these change into the future.

Rainfall associated with the occurrence and co-occurrence of events is presented in Figure 74 for ERA5 (bottom row), and 24 GCMs over the common period of 1961 to 2014 as a percentile of the full rainfall distribution for the region (Figure 74a). For example, considering the ERA5 reanalysis row and the positive IOD event column, this is interpreted as that during a positive IOD event the rainfall typically falls into the 73rd percentile; then for a negative IOD event, it falls into the 24th percentile. Largest rainfall percentiles are evident during co-occurring events of +IOD and +ENSO (81st percentile) and lowest rainfall percentile during the co-occurring events of -IOD and -ENSO (17th percentile). Seasons where pure ENSO or IOD events occur are common, whereas seasons where both events co-occur are rare (Figure 74b).

Rainfall distribution anomalies are statistically significant at the $p < 0.01$ level during +IOD, -IOD, +IOD/+ENSO, and -IOD/-ENSO, Figure 74a has been masked to only show statistically significant anomalies at $p < 0.01$. Positive IOD events are associated with higher rainfall percentiles (wetter), -IOD events with lower rainfall percentiles (drier), co-occurring +IOD and -ENSO (wetter) and -IOD and -ENSO (drier). However, the frequency of the co-occurring modes is low compared to IOD-only events.

Therefore, based on the ERA5 data the conclusions are that (1) the IOD is the major driver of statistically significant rainfall anomalies in the catchment region and (2) that although co-occurrence of events is comparatively very rare, a co-occurring ENSO event of the same sign as an IOD event amplifies the rainfall anomaly.

a)



b)

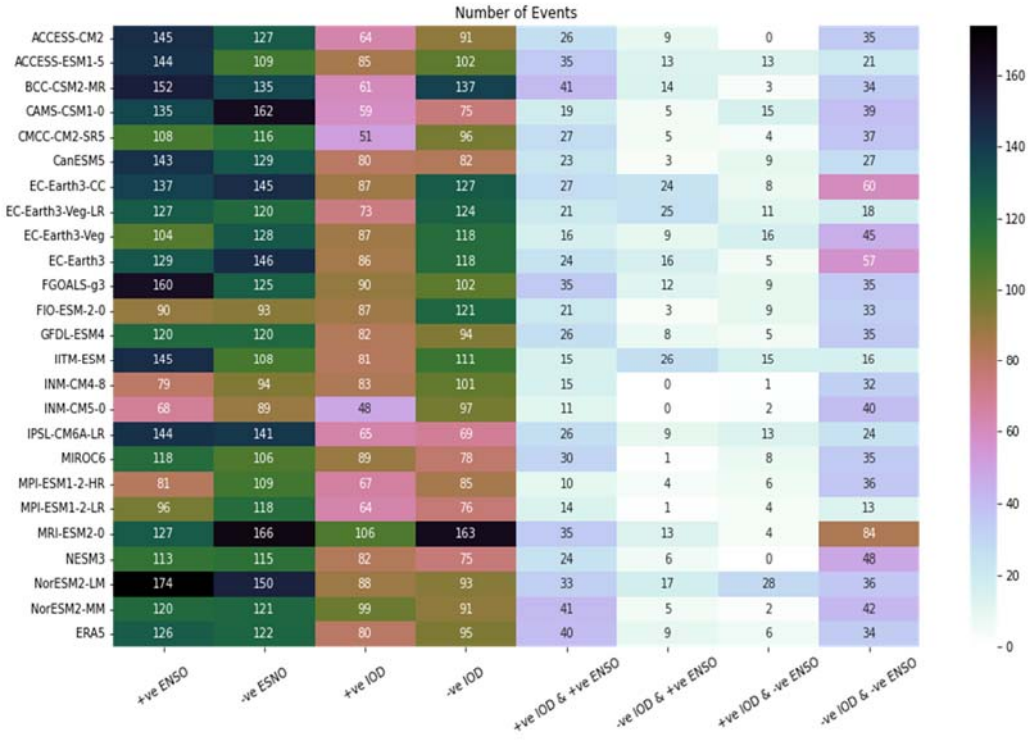


Figure 74: (a) Rainfall associated with the occurrence and co-occurrence of events as a percentile of the full rainfall distribution for the region over the common period of 1961 to 2014. Only statistically significant values at the p<0.01 level are shown. The ERA5 results are in the bottom row and the 24 GCMs in the rows above. (b) The frequency of events represented in the ERA5 data and GCMs.

There is a large heterogeneity in the representation of the rainfall distribution anomalies by the 24 GCMs. For instance, in some modes a larger number of models had rainfall anomaly percentiles above the ERA percentile, e.g. -ENSO and -IOD.

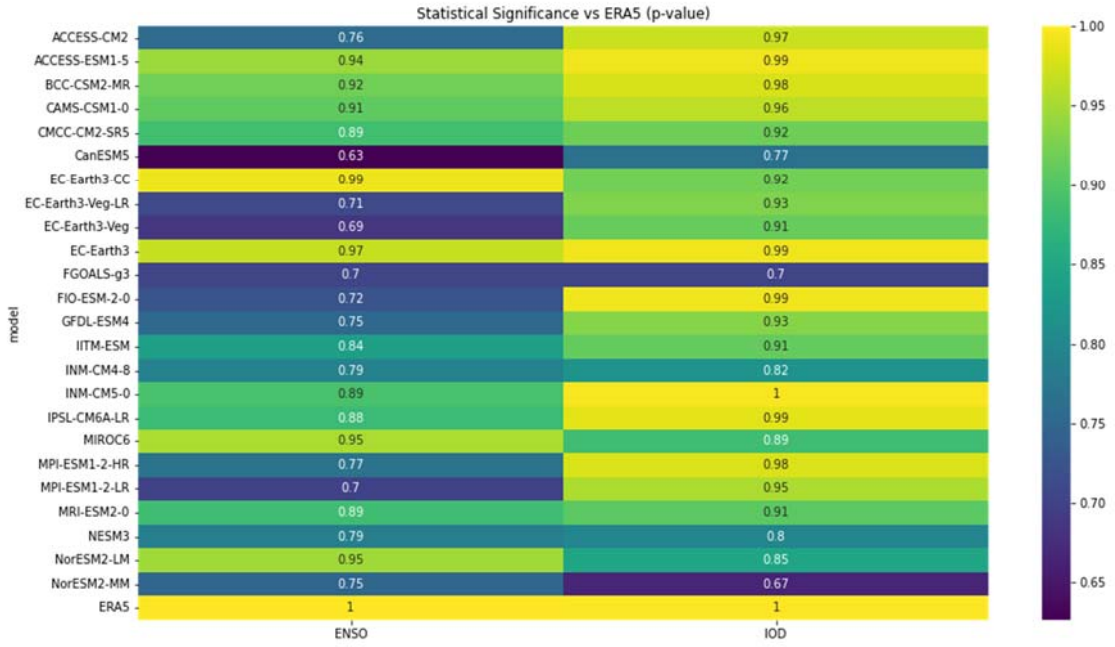


Figure 75: Statistical significance of the difference between the GCM ENSO and IOD frequency in comparison to the ERA5 frequency. All values less than the p-value of 0.01 according to an independent 2-sample t-test are considered significantly different from the ERA5 results and those above, significantly similar to the ERA5 distribution.

In terms of how well the GCMs replicate the ERA5 results. The frequency of both ENSO and IOD results is largely similar across the models. However, when considering Figure 75, which accounts for the statistical likelihood of the simulated ENSO and IOD indices coming from the same distribution as that simulated by ERA5 reanalysis, all but four CMIP6 models do a better job simulating IOD than ENSO.

Thus, it has been demonstrated that the IOD is the most important forcing of rainfall in the catchment region, and that there is a measure of confidence in the ability of the GCMs to reproduce spatial and statistical characteristics of the IOD. It has also been shown this is less true for ENSO. The impact of anthropogenic heating under the forcing scenario SSP245 for the period 2030-2059 on respective event-forced rainfall is considered hereafter.

Fewer GCMs simulate a statistically significant percentile of the full rainfall distribution to be associated with the +IOD phase in the future (9 vs 11 for the historical period), the same is true for the -IOD phase (10 vs 14) (Figure 76). Little change is evident for co-occurring +IOD/+ENSO (10 vs 9) and -IOD/-ENSO phases (11 vs 12).

The rainfall percentile associated with a +IOD event in the future is projected to decrease significantly from the 67th percentile on average during the historical period to the 61st percentile during the 2030-2059 period. This suggests that rainfall associated with +IOD will change from wetter than normal

during the historical period to more normal in the medium future. There are only small differences in percentile changes of other event modes.

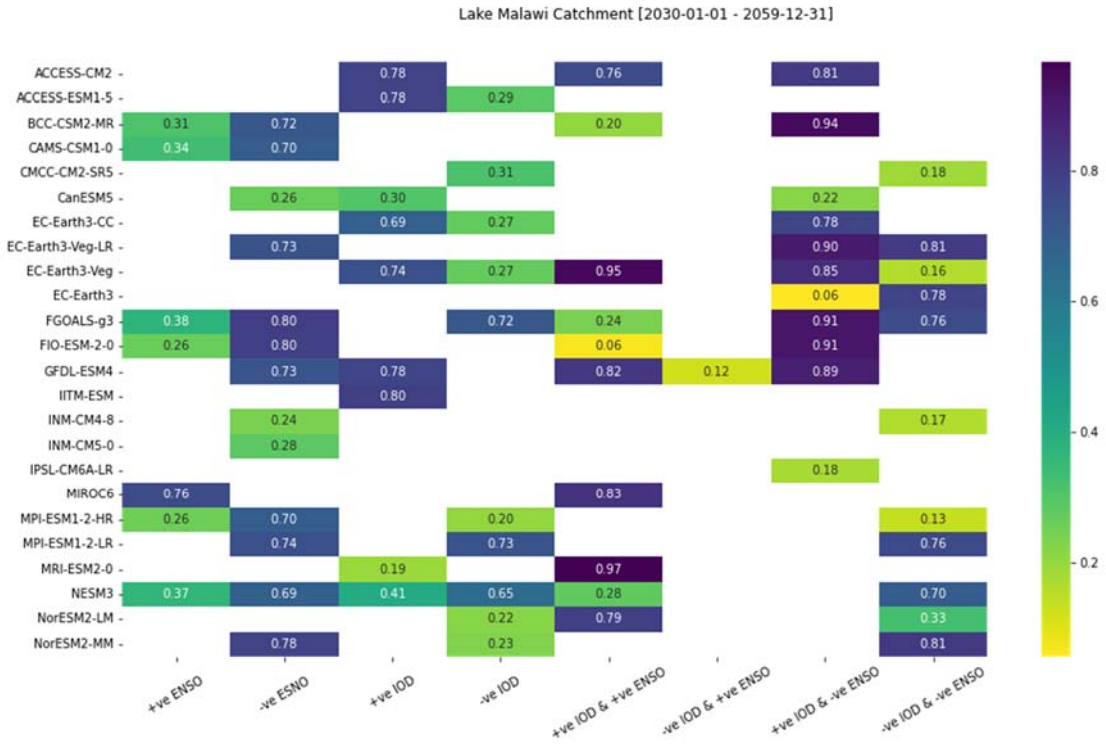


Figure 76: Projected rainfall associated with the occurrence and co-occurrence of events as a percentile of the full rainfall distribution for the region over the period 2030-2059 under the SSP245 scenario. Only statistically significant values at the p<0.01 level are show.

4.4.2 *Projected changes in synoptic drivers of rainfall over Malawi*

The WP3 study identified two large-scale moisture feeds into the Malawi region: (1) a north westerly flow during which moisture is advected to the region over Central Africa from the tropical Atlantic and is associated with more homogeneous regional scale rainfall, and (2) an easterly flow that advects moisture from tropical Indian Ocean and is associated with more local rainfall. The analysis presented here investigates how these large-scale moisture fluxes may change as a result of anthropogenic warming and the impact this would have on rainfall in the region.

It does so through an analysis of atmospheric processes derived from prognostic variables that are produced by climate models. Prognostic variables are those governed by prognostic equations that solve for conservation of mass, momentum, and thermodynamic energy through integrations in time and include variables like geopotential height, 3-dimensional fields such as wind, temperature and humidity. Diagnostic variables, on the other hand, are derived variables that are computed from prognostic variables and other external parameters e.g., rainfall, 2-meter temperature and humidity fields and winds at 10 meters amongst others.

Through assessing projected changes in large-scale prognostic fields that establish the environment for the regional daily weather response through smaller scale circulations associated with them (e.g. convective systems, land and sea breezes), instead of the simulated rainfall field (a diagnostic variable notoriously difficult to simulate as it is a function of many other variables and parameterizations), a more reliable and defensible rationale can be developed to explain potential changes in the nature of projected rainfall over the region.

Self-organizing maps (SOMs) have been used to identify characteristic atmospheric states and their influence on rainfall over Malawi. This is achieved through the lens of moisture advection, and to quantify projected changes in the frequency of occurrence of atmospheric states that may be associated with a particular moisture advection regime and to discern whether there is a seasonal shift in these patterns.

4.4.2.1 Data and method

Relating large scale characteristics to local scale responses requires the reduction of many variables into a smaller set of data that still represent the original information. For example, precipitation is a function of many variables, e.g. pressure, temperature, humidity etc. and interactions between these variables have to be preserved when relating synoptic characteristics to the local scale precipitation response. Self-organizing maps (SOMs) use multivariate atmospheric data to produce several generalized weather circulations over a chosen period. This is achieved through a non-linear projection of the probability density function of high-dimensional input data onto a two-dimensional array of nodes while spanning the full continuum of the data space.

Through an iterative training process, the SOM identifies a prior determined number of nodes within the multivariate climate data space such that the nodal distribution represents the observed distribution, providing a means for data to be generalized into a number of arch-types, or a 'trained'

SOM. The data used to develop the trained SOM maps to the best matching node (here large-scale archetypal circulation patterns), so the frequency each archetypal circulation is mapped to can be quantified. Furthermore, data from other climate models such as projection data from GCMs, can be mapped to the trained SOM and the differences in frequency mapping computed to quantify differences between the different models. Here, the differences in the frequency mappings between the trained SOM and projection data inform on changes in the large-scale circulation states which in turn inform on how rainfall in the region may change.

Daily synoptic atmospheric data from ERA5 reanalysis at 700hPa between 1981-2020 is used to develop the trained SOM, which is considered to represent discrete 'observed' states of the atmosphere. From the work in WP3, three variables are selected to represent moisture transport into the region, namely daily fields of specific humidity and u- and v-wind components at 700 hPa. This vertical level is used as below 850hPa high topography disrupts the variable field and regrettably, some GCMs do not provide data at 800hPa. Twelve CMIP6 GCMs are analyzed - all data has been regridded to a common 100 km resolution using a bilinear spline. Data has been standardized over the period of 1985-2020 to preserve local gradients in each field and transformed through a Principle Component Analysis. The first 8 principal components are retained for the SOM training procedure (see below).

The GCM historical period selected for comparison with the trained SOM is 1985-2020 and the projection period selected is 2031-2060 to represent what the IPCC refers to as the near- and mid-future. Only one scenario, SSP245, has been considered to represent the scenario global emissions are currently following.

4.4.2.2 Results

Archetypal moisture flux states (nodes) produced by the SOM are shown in Figure 77. Over the study region easterly, southeasterly, and northeasterly winds are present in most of the nodes except in the bottom right of the SOM where westerly winds occur (nodes 15 and 20). The top row of the SOM (nodes 1-5) is associated with the passage of mid-latitude cyclones and easterly winds over the catchment area, and relatively low specific humidity values over the study region. SOM states representing an anticyclonic circulation over South Africa are in the bottom left and last row of the SOM. This circulation state is associated with a south easterly advection of relatively dry air into the study region. Only nodes 15 and 20 are associated with westerly moisture advection, which in D3.2 was shown to contribute ~30% to total annual rainfall, and have a larger regional average (as opposed to local averages) than rainfall associated with the easterly flows.

Rainfall associated with each circulation state is shown in Figure 78. Large-scale, mean area averaged rainfall over the Lake Malawi domain is associated with the bottom and right of the SOM, with maximum values in nodes 15 and 19, again demonstrating that westerly flow brings large scale rainfall (Figure 78a). However, easterly flows associated with nodes 10, 14, 17 and 18 also account for a large proportion of large-scale rainfall. Localized rainfall maxima also maps to the right and bottom nodes of the SOM (Figure 78b). This indicates that local maxima are associated with a wider range of synoptic states, however, nodes 15 and 20 again are associated with the highest local maxima, indicating they are important moisture flux states with respect to both area-wide rainfall and local maxima.

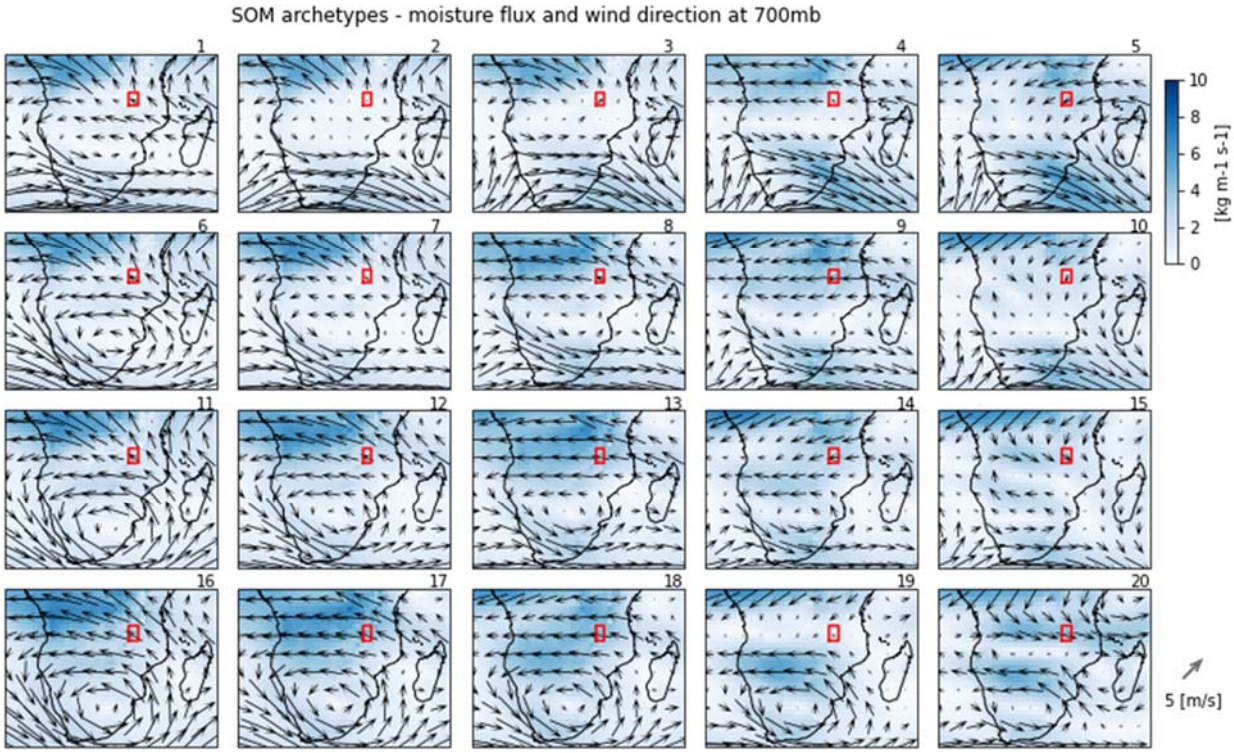


Figure 77: Archetypal synoptic scale moisture flux states produced by the SOM based on the ERA5 reanalysis data over the period 1985-2020. Blue shading shows specific humidity and arrows are wind vectors. The red block represents the Lake Malawi river catchment.

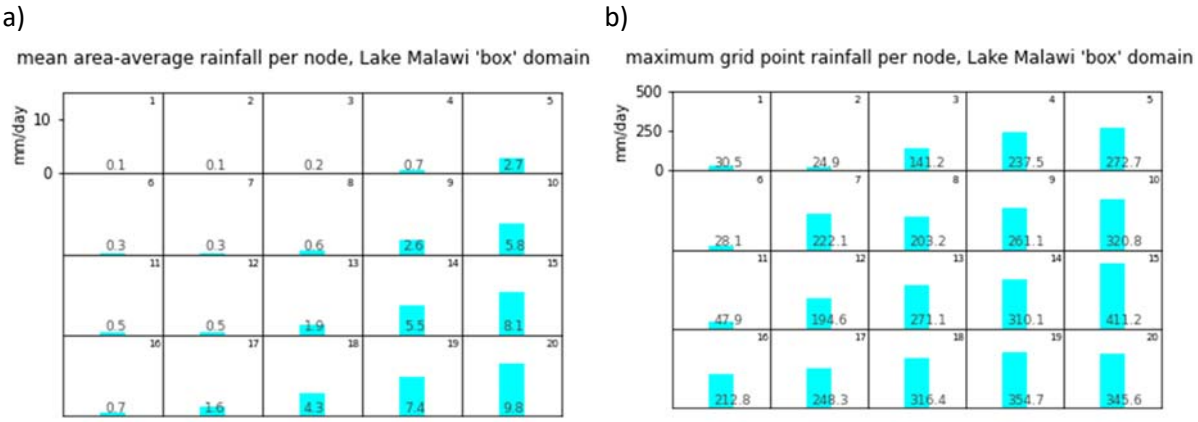
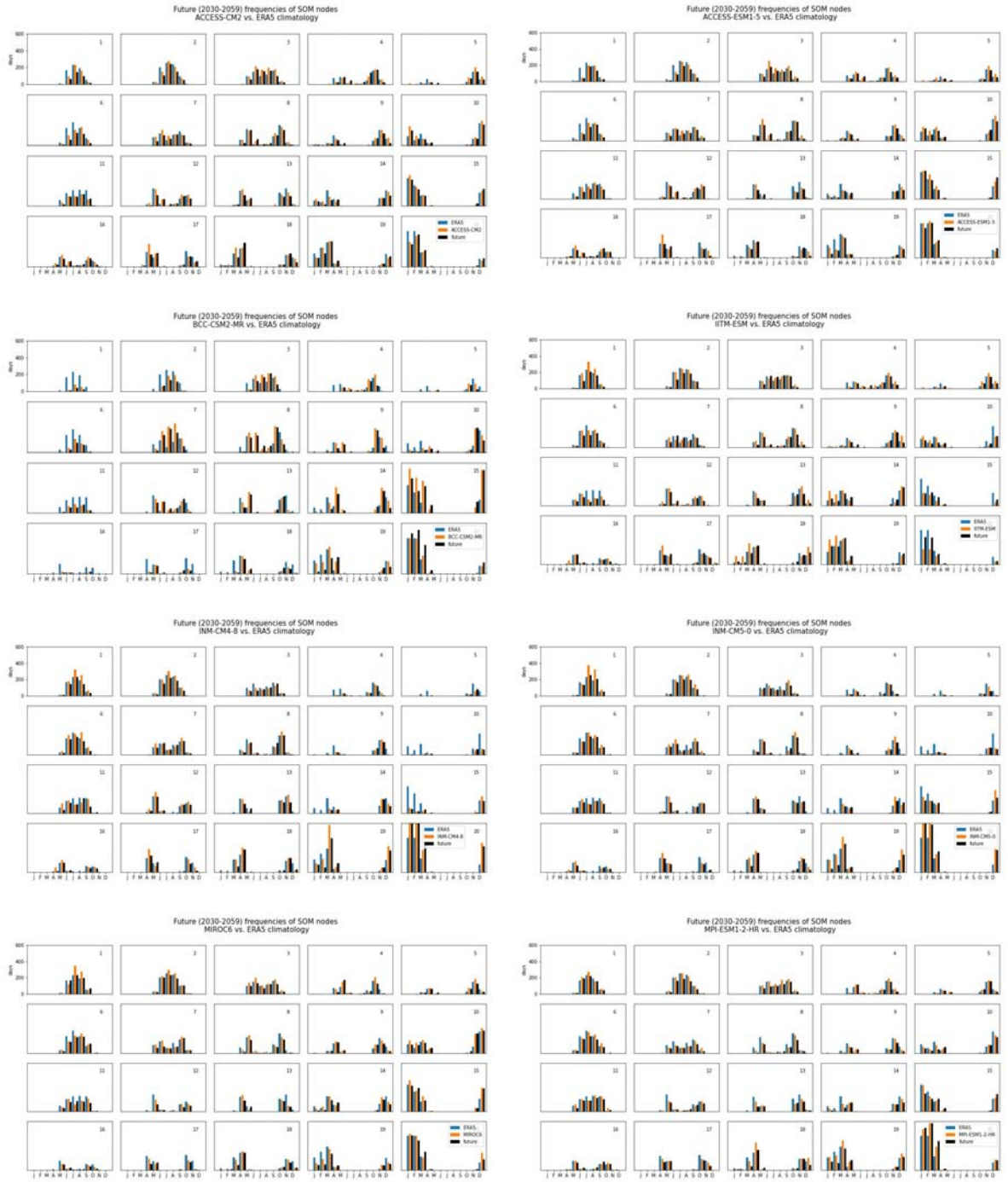


Figure 78: Rainfall associated with each SOM node in mm/day. Rainfall area averaged over the catchment is shown in (a) and the maximum grid point value in the Lake Malawi 'box' is shown in (b) to quantify extreme localized rainfall.

Monthly mapping of the ERA5 and GCM historical and scenario data to the trained SOM is presented in Figure 79, as the number of days per month that map to each node. From the ERA5 mapping (blue bar), winter circulation states lie to the left of the SOM and, as one moves across to the right, this transitions to shoulder seasons and summer circulation states. Nodes 15 and 20 are associated with DJFM and the core of the rainy season over the catchment. The majority of GCMs appear to have

mappings that replicate that of ERA5 over the historical period (orange bar), although IITM-ESM, INM-CM5-0, INM-CM4-8 and BCC-CSM2-MR show considerable disparities in the rain-generating nodes 15 and 20.

In terms of future projections (2030-2059), and excluding the four poorly performing models above, monthly mappings to rainfall producing nodes 15, 19 and 20, do not show any substantial change compared to the historical period. In other parts of the SOM fewer mappings in the projected period are noted in nodes 2 and 3 by 7 GCMs, and in nodes 6 and 1,1 by 5 GCM (however, these nodes are not associated with rainfall).



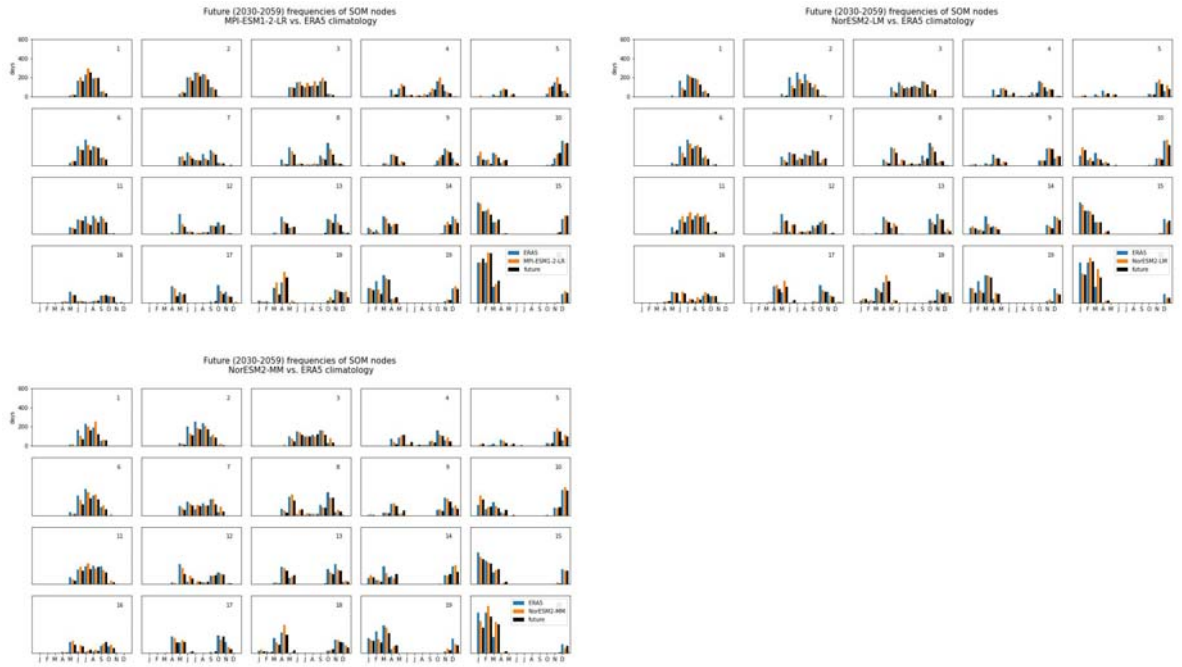


Figure 79: Monthly mapping of the ERA5 reanalysis data (blue), GCM historical data (orange) and GCM projection data (black) to each node as days per month. Each map is for one GCM.

4.5 Analysis of drought-related variables in southern Africa

Generally, across southern Africa it has been shown that model agreement of the Coupled Model Intercomparison Project Phase Six (CMIP6), Coordinated Regional Downscaling Experiment (CORDEX) and CORDEX-core ensembles is strong in terms of the general pattern of projected decreases in rainfall (Dosio et al. 2021). However, there is less agreement over Mozambique and the eastern escarpment areas of South Africa, where some models do not project general reductions in rainfall totals, but rather rainfall increases. This model uncertainty persists even under high levels of global warming (Lee et al., 2021), suggesting that it is the result of structural rather than internal uncertainty. Uncertainties in cumulus parameterization are often suggested as the source of this uncertainty, and indeed models exhibit substantial biases in their depiction of rainfall totals over the eastern escarpment regions (Dedekind et al., 2016).

In this section of D4.1, the focus is on exploring in more detail the weather extremes associated with the main message of change projected for southern Africa (namely of the region becoming generally drier and drastically warmer). More specifically, the attention is on the combined effect of changes in heat-wave frequencies and a general trend of drying, in the context of potentially devastating impacts on agriculture under low mitigation climate change futures. This analysis has been conducted across the southern African region, covering all the FOCUS case-study areas.

4.5.1 Data and methods

Daily rainfall, average temperature, minimum and maximum temperature, relative humidity, and surface wind speed data have been obtained for the CMIP6 ensemble (1961-2100; Tebaldi et al., 2021) of global climate models (GCM) as well as from the CORDEX-core ensemble (2005-2099; Gnitou et al., 2021) of regional climate models (RCM). These six surface variables are essential for the calculation of the Keetch-Byram drought index that is key to the analysis undertaken in this section. For the CMIP6 ensemble, six models have the mentioned six surface variables under the low mitigation scenario SSP5-8.5 (Socio-economic Pathway 5-8.5). On the other hand, nine models that completed the CORDEX-core simulations have the six required surface variables under the equivalent low-mitigation scenario, RCP8.5 (Representative Concentration Pathway 8.5). Note that CORDEX-core simulations used in this analysis are downscaled from the GCMs of the Coupled Model Intercomparison Project Phase Five (CMIP5).

The relatively low resolution CMIP6 GCM data has been interpolated to a common 1° latitude-longitude grid, to allow a model-intercomparison of the projected climate change futures. Hence, the CORDEX-core model simulations, which have a resolution close to 0.22°, have been interpolated to the same 1° grid. For the CMIP6 data, the baseline period has been selected to be 1961-1980, and the future period, 2080-2099 (the long-term future in IPCC terminology). These periods have been selected to allow an optimal assessment of the climate change signal. In the case of the CORDEX-core data, the baseline period has been selected to be 2005-2014, with the future period being also 2080-2099.

In addition to the CMIP6 and CORDEX-core ensembles, an ensemble of detailed projections of future climate change generated from the conformal-cubic atmospheric model (CCAM), available at 0.1° resolution over southern Africa has been used (see FOCUS-Africa D3.1 for a detailed description of the experimental design). In short, CCAM was used to downscale six GCMs of CMIP5, under RCP8.5, for the period 1961-2099. In the case of CCAM, the period 1961-1980 has been selected as the baseline period, and 2080-2099, as the long-term future. This CCAM ensemble also contains the six key variables relevant to the calculation of the Keetch-Byram drought index.

Two extreme weather-event definitions have been employed in the following analysis. The first is the World Meteorological Organization (WMO) definition for heat-waves, as events when the maximum temperature at a specific location exceeds the average maximum temperature of the warmest month of the year by 5 °C, for a period of at least 3 days (Engelbrecht et al., 2015). The second is the Keetch-Byram drought index, D, which is defined in terms of a daily drought factor, dQ:

$$dQ = \frac{(203.2 - Q)[0.968 \times e^{(0.0875T + 1.5552)} - 8.30]}{[1 + 10.88 \times e^{(-0.001736R)}]}$$

Here R, is the mean annual precipitation (mm) and Q (mm), is the soil-moisture deficiency that results from the interaction between rainfall and evaporation. Once Q has been updated by dQ, the drought index is calculated from the equation:

$$D = \frac{10Q}{203.2}$$

Note that D ranges from 0 to 10, where D=10 indicates completely dried out soil and vegetation (Keetch and Byram, 1968, Engelbrecht et al., 2015).

4.5.2 Results

4.5.2.1 Projected climate change in the CMIP6 ensemble

4.5.2.1.1 Rainfall

The six ensemble members considered (with daily data available), display a future general pattern of drying over southern Africa (for the period 2080-2099 relative to the baseline period 1961-1980, Figure 88). Four out six members project general rainfall increases over the eastern escarpment of South Africa. The CMIP6 models also envisage a pronounced wetting over eastern Africa, extending to central tropical Africa and western Africa (a pattern that is not evident from the CORDEX-core projections).

4.5.2.1.2 Temperature

Consistent with the CCAM and CORDEX-core ensembles, the CMIP6 projects drastic warming over the western and southern interior regions of southern Africa (Figure 89), which in two ensemble members even exceed 6°C for the period 2080-2099 (relative to the baseline period of 1961-1980).

4.5.2.1.3 Drought Index

Despite the extensive rainfall buildup that some CMIP6 members project for the eastern escarpment of southern Africa. domain, the Keetch-Byram drought index also rises, suggesting an overwhelming effect of intense evaporation on soil-moisture (Figure 90). That said, CMIP6 ensemble also foresees soil-moisture increases in tropical Africa.

4.5.2.1.4 Heat-wave days

Drastic increases in the number of heat-wave days are projected by the CMIP6 ensemble across the interior regions of southern Africa, consistent with the CCAM and CORDEX-core ensembles (Figure 91). Most of the projections also indicate a pronounced growth in heat-wave days in western and central tropical Africa, consistent with the CORDEX-core ensemble.

4.5.2.2 Projected future climate change in the CCAM ensemble

4.5.2.2.1 Rainfall

A generally drier future is projected for southern Africa in the long-term (2080-2099) in the CCAM ensemble, relative to the present-day baseline (1961-1980, Figure 80). The pattern of drying is particularly intense for the winter rainfall region of the southwestern Cape, across all the ensemble members. Extensive rainfall reductions are also projected for northeastern South Africa, and westwards into the Limpopo River Valley, Botswana and the FOCUS-Africa area of interest of the northwest province in South Africa. There are two exceptions to this general drying behavior: over coastal KwaZulu-Natal in South Africa, the Lesotho Drakensberg and adjacent interior of South Africa to the west, several ensemble members are indicative of rainfall increases. These general rise in rainfall totals occur in association with a proliferation in the frequency of intense rainfall events (not shown here). It can be deduced that these growths are the result of an increase in convective thunderstorms. The other exception is southern Mozambique, where some ensemble members (not all) show an increase in rainfall. This could be likely the result of a more frequent landfall of tropical lows and cyclones, and/or, landfalling lows and cyclones causing more rainfall than in the past.

4.5.2.2.2 Temperature

The projections of temperature changes under low mitigation scenarios, for the period 2080-2099 (relative to the 1961-1980 baseline), are suggestive of solid rises in the annual average temperature, likely ranging between 4°C and 7°C (
Figure **81**). Consistent with trends that have already been detected, the strongest warming is projected for Botswana, extending in some ensemble members to the northwest province of South Africa.

4.5.2.2.3 Drought index

The Keetch-Byram drought index is indicative of general reductions in soil-moisture in southern Africa in the substantially warmer long-term future (Figure 82). Whereas for Lesotho and southern Mozambique there is model disagreement in terms of rainfall changes within the ensemble, all members display reductions in soil-moisture even for these regions. This is the consequence of enhanced evapotranspiration in a substantially warmer long-term future.

4.5.2.2.4 Heat-wave days

The projected increase in the annual number of heat-wave days in the long-term is more than 60 days per year across Botswana and the North West province in South Africa, with somewhat smaller growths elsewhere in the southern African interior and along the coastal areas (Figure 83). It is noteworthy that this key hazard to agriculture and human health in southern Africa is projected to be compounded by general reductions in soil-moisture and more frequently occurring droughts (Figure 82).

4.5.2.3 Projected climate change in the CORDEX-core ensemble

4.5.2.3.1 Rainfall

The CORDEX-core ensemble envisages a general pattern of drying across southern Africa (for the period 2080-2099 relative to 2006-2015), but with most ensemble members indicative of rainfall increases over Tanzania (Figure 84). Conversely, there is model disagreement in terms of the projected climate change signal over the eastern escarpment of South Africa, with some ensembles suggestive of sizeable rainfall increases, whilst others project intense drying. There is similarly model disagreement in terms of the future rainfall signal over northern Madagascar, with half of the members showing rainfall increases and the other half indicating decreases. This implies that the models have a pronouncedly different representation of future tropical low and cyclone tracks. Most of the members display a general drying over Mozambique, with only one of the ensemble members denoting rainfall increases over the southern part of the country (a signal present in several of the CCAM downscalings).

4.5.2.3.2 Temperature

In this variable, three of the six ensemble members indicate an extreme warming of more than 4°C for the period 2080-2099 relative to the 2006-2015 baseline (with somewhat reduced warming projected by the remainder of the ensemble, Figure 85). As in the CCAM ensemble, the projected warming is strong over Botswana, with some ensemble members extending the area of higher warming into the northwest province of South Africa (a FOCUS-Africa area of interest).

4.5.2.3.3 Drought Index

When considering future changes in soil-moisture, as approximated by the Keetch-Byram drought index, the CORDEX-core ensemble shows robust model agreement on decreases across the southern African region (Figure 86). As in the case of the CCAM ensemble, this holds true even in areas where there is model disagreement in terms of the rainfall signal. This can be attributed to the substantial surface temperature growths, which can be directly linked to an intensification of evapotranspiration processes.

4.5.2.3.4 Heat-wave days

The CORDEX-core ensemble foresees drastic increases in the annual number of heat-wave days across the western and central interior of southern Africa, with several ensemble members indicating growths of more than 40 of such days per year in some areas (Figure 87). Similarly, considerable rises are also projected for western tropical Africa, whereas smaller are indicated for eastern Africa (including Tanzania and Mozambique). This substantial expansion in heat-waves, in combination with generally drier conditions and more frequent agricultural drought (Figure 86) may be regarded as an important climate risk southern Africa will have to face in the future.

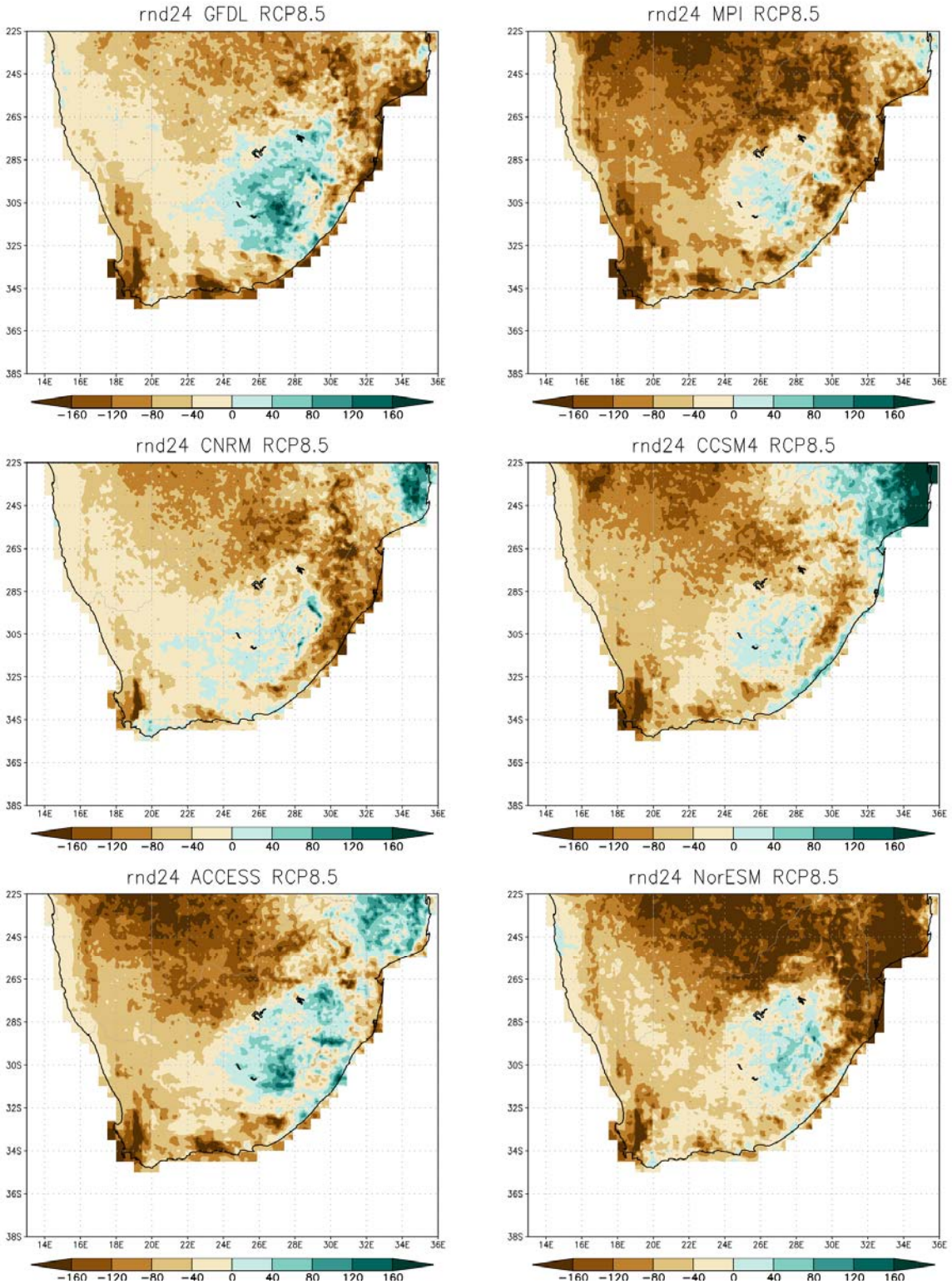


Figure 80: CCAM-ensemble projected change in annual rainfall (mm) over southern Africa at 8 km resolution, for the time-slab 2080-2099 relative to 1961-1980, under low mitigation.

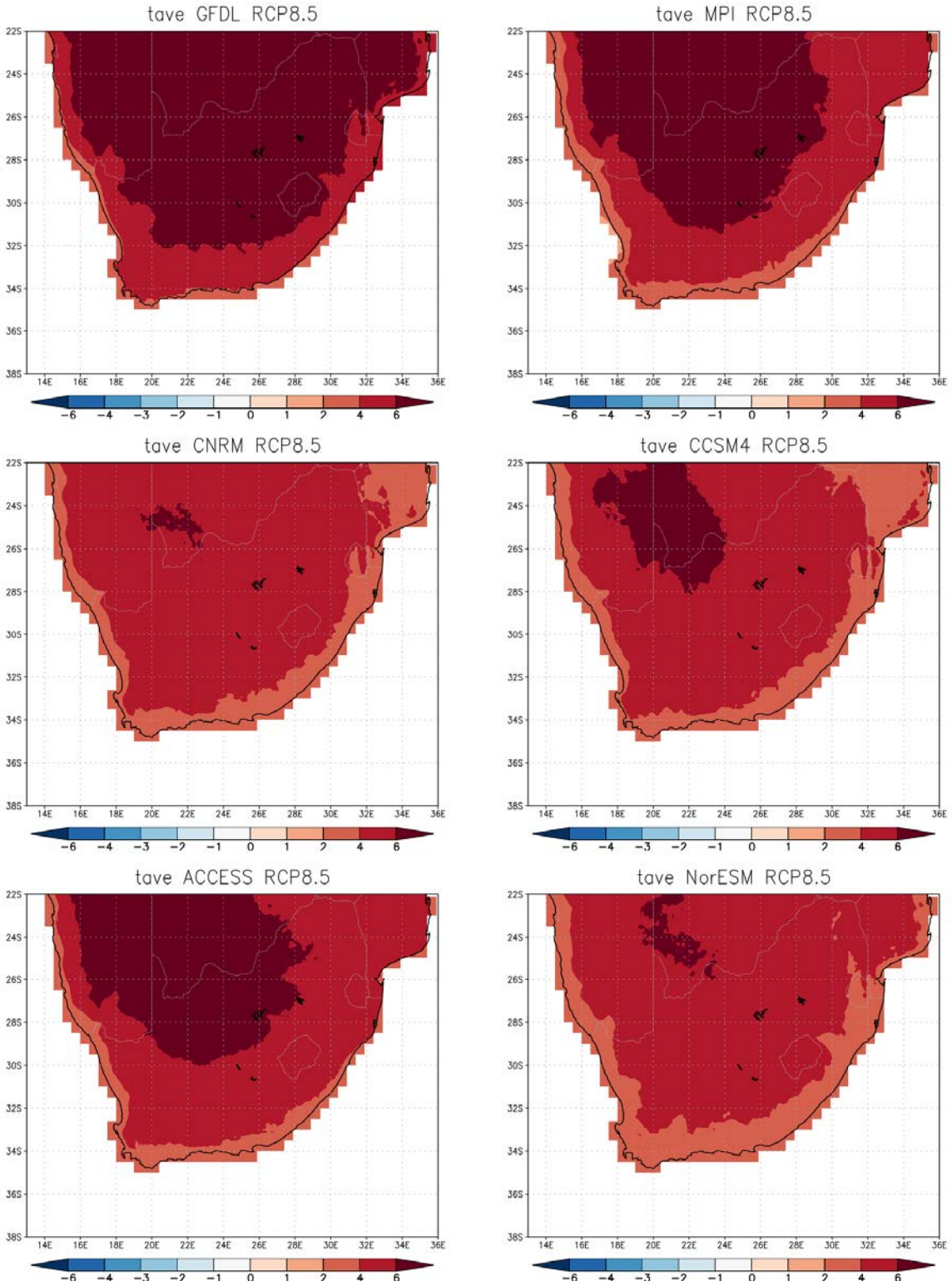


Figure 81: CCAM-ensemble projected change in the annual average temperature (°C) over southern Africa at 8 km resolution, for the time-slab 2080-2099 relative to 1961-1980, under low mitigation.

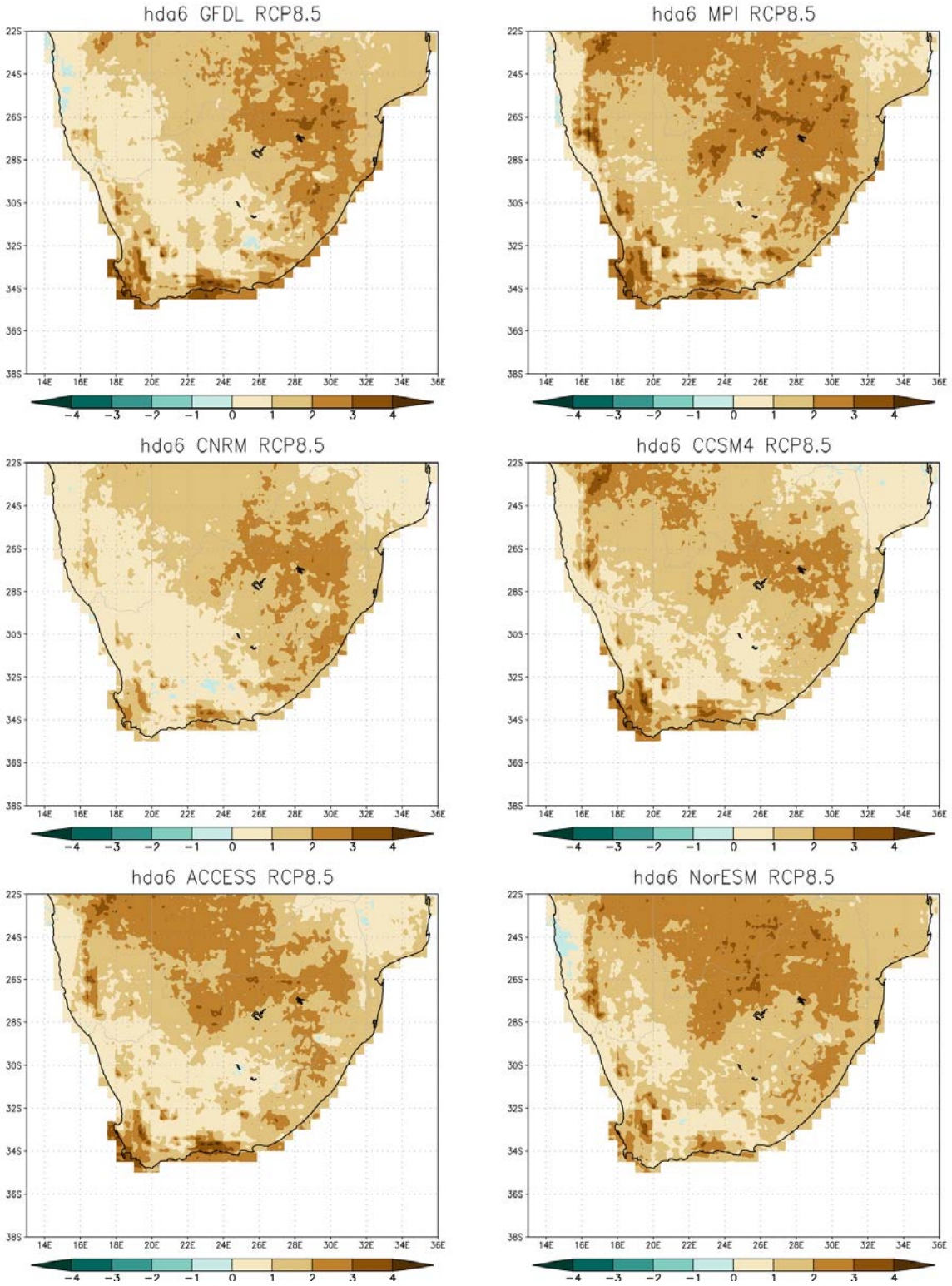


Figure 82: CCAM-ensemble projected change in the Keetch-Byram drought index over southern Africa at 8 km resolution, for the time-slab 2080-2099 relative to 1961-1980, under low mitigation.

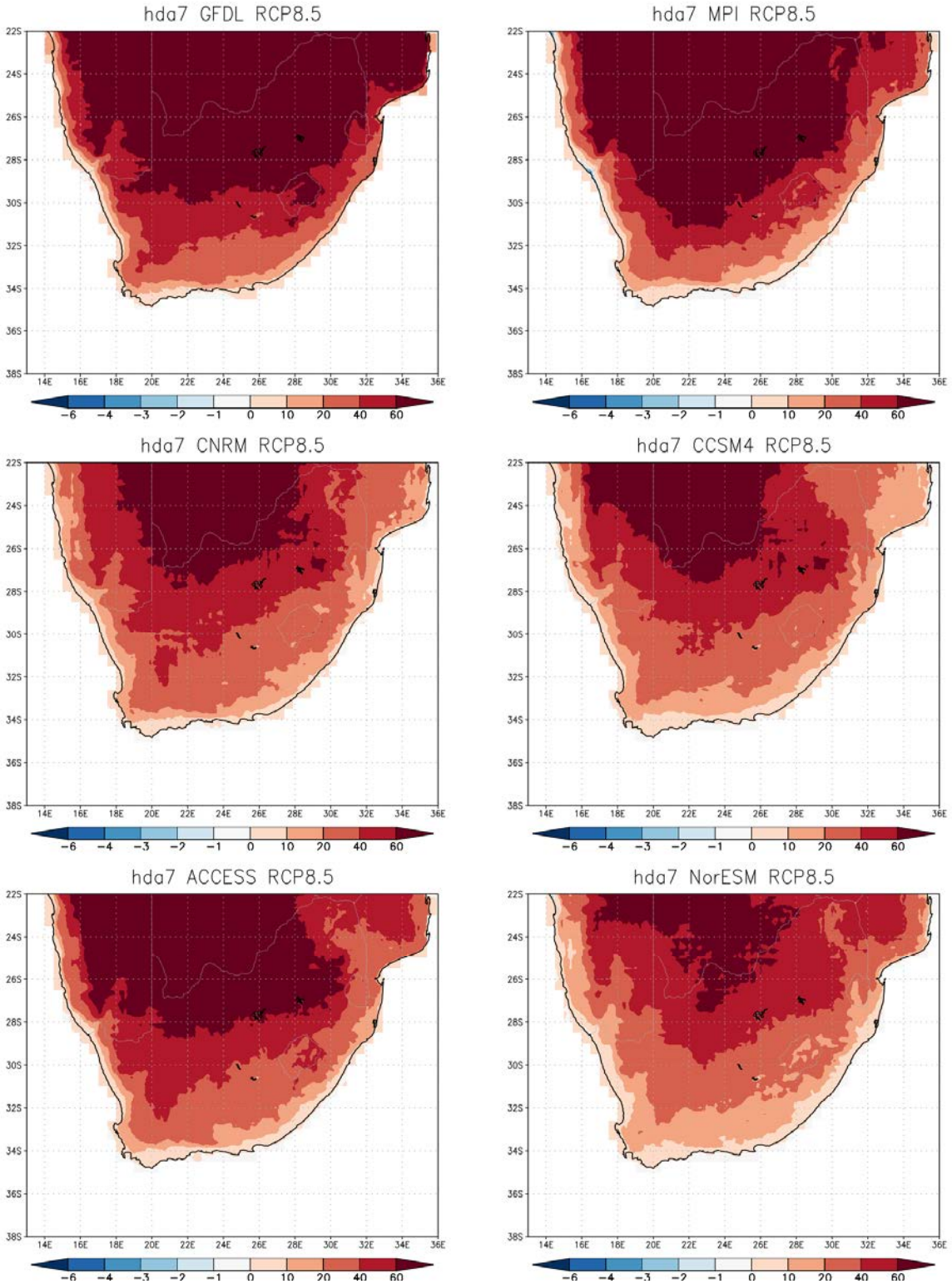


Figure 83: CCAM-ensemble projected change in the number of heat-wave days over southern Africa at 8 km resolution, for the time-slab 2080-2099 relative to 1961-1980, under low mitigation.

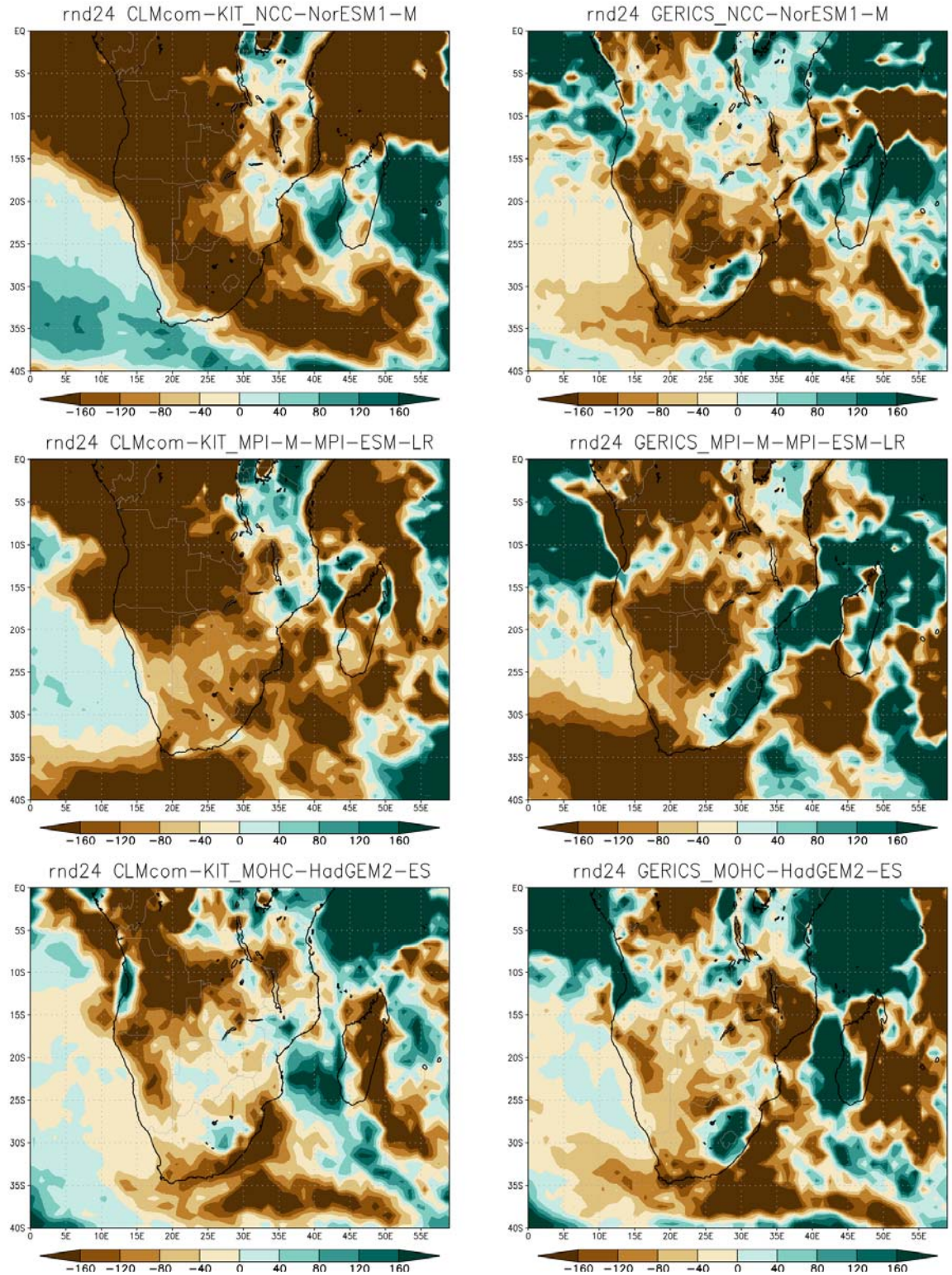


Figure 84: CORDEX-core ensemble projected change in annual rainfall (mm) over southern Africa, for the time-slab 2080-2099 relative to 2006-2015, under low mitigation.

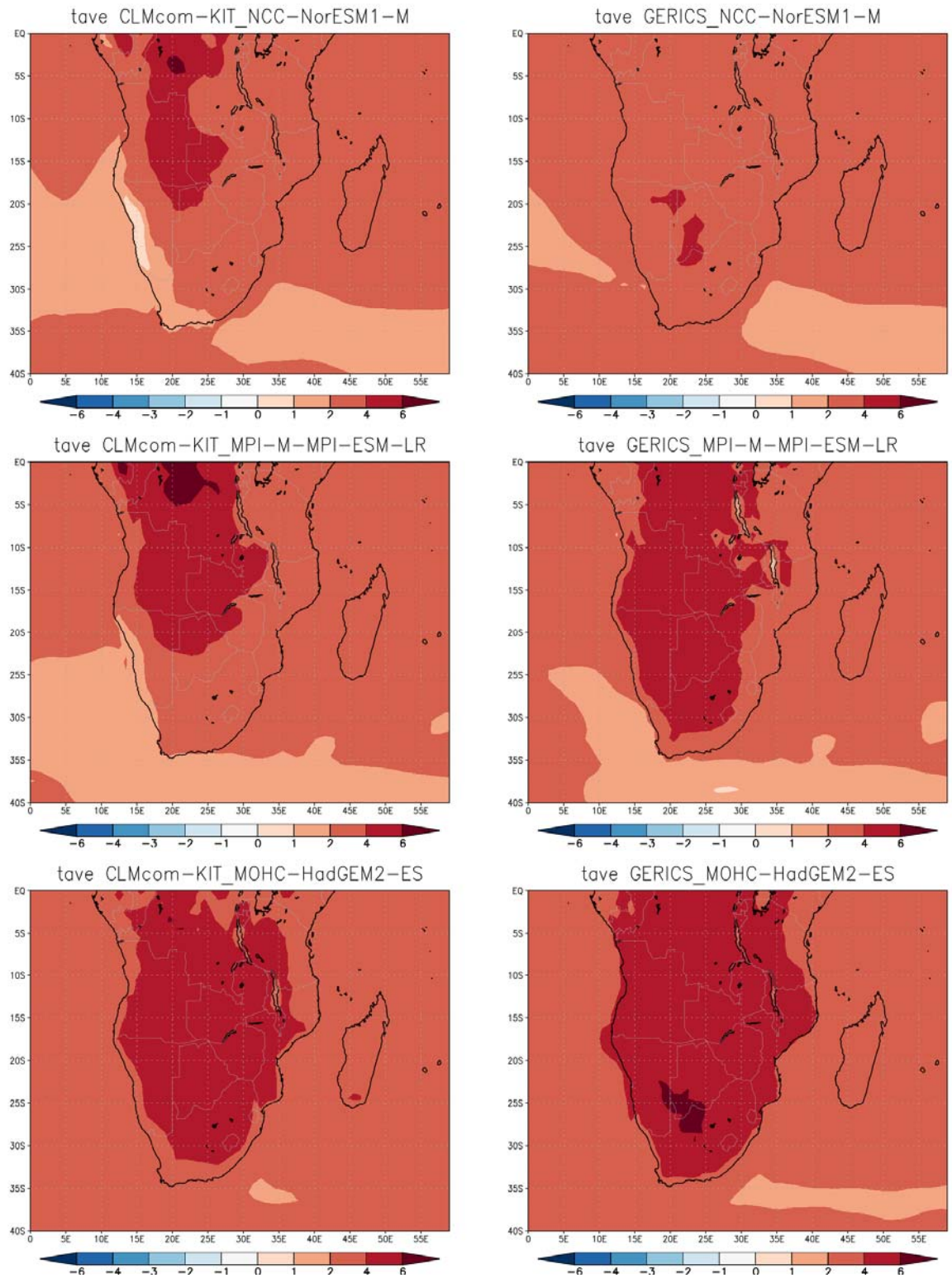


Figure 85: CORDEX-core ensemble projected change in the annual average temperature (°C) over southern Africa, for the time-slab 2080-2099 relative to 2006-2015, under low mitigation.

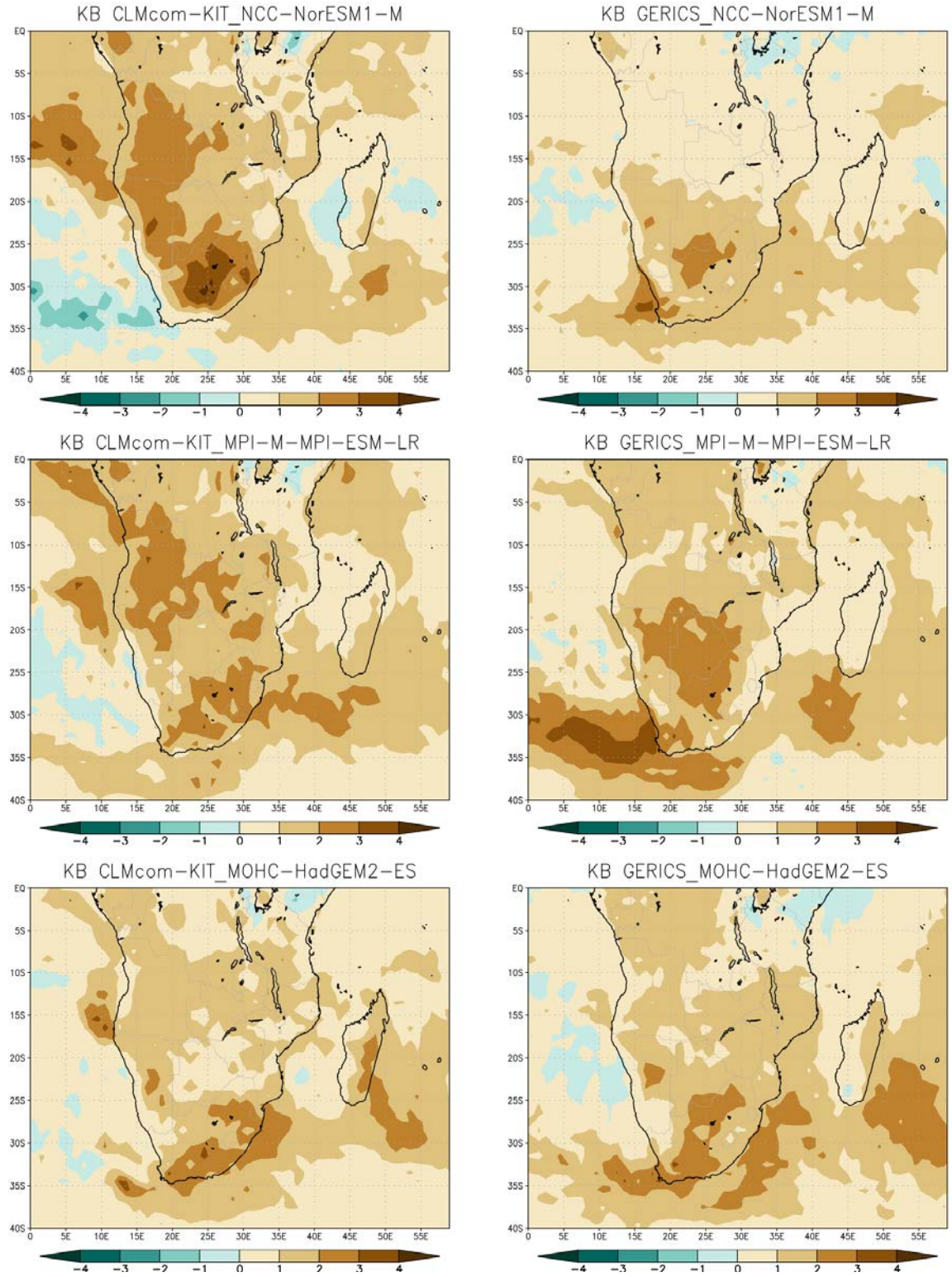


Figure 86: CORDEX-core ensemble projected change in the Keetch-Byram drought index over southern Africa, for the time-slab 2080-2099 relative to 2006-2015, under low mitigation.

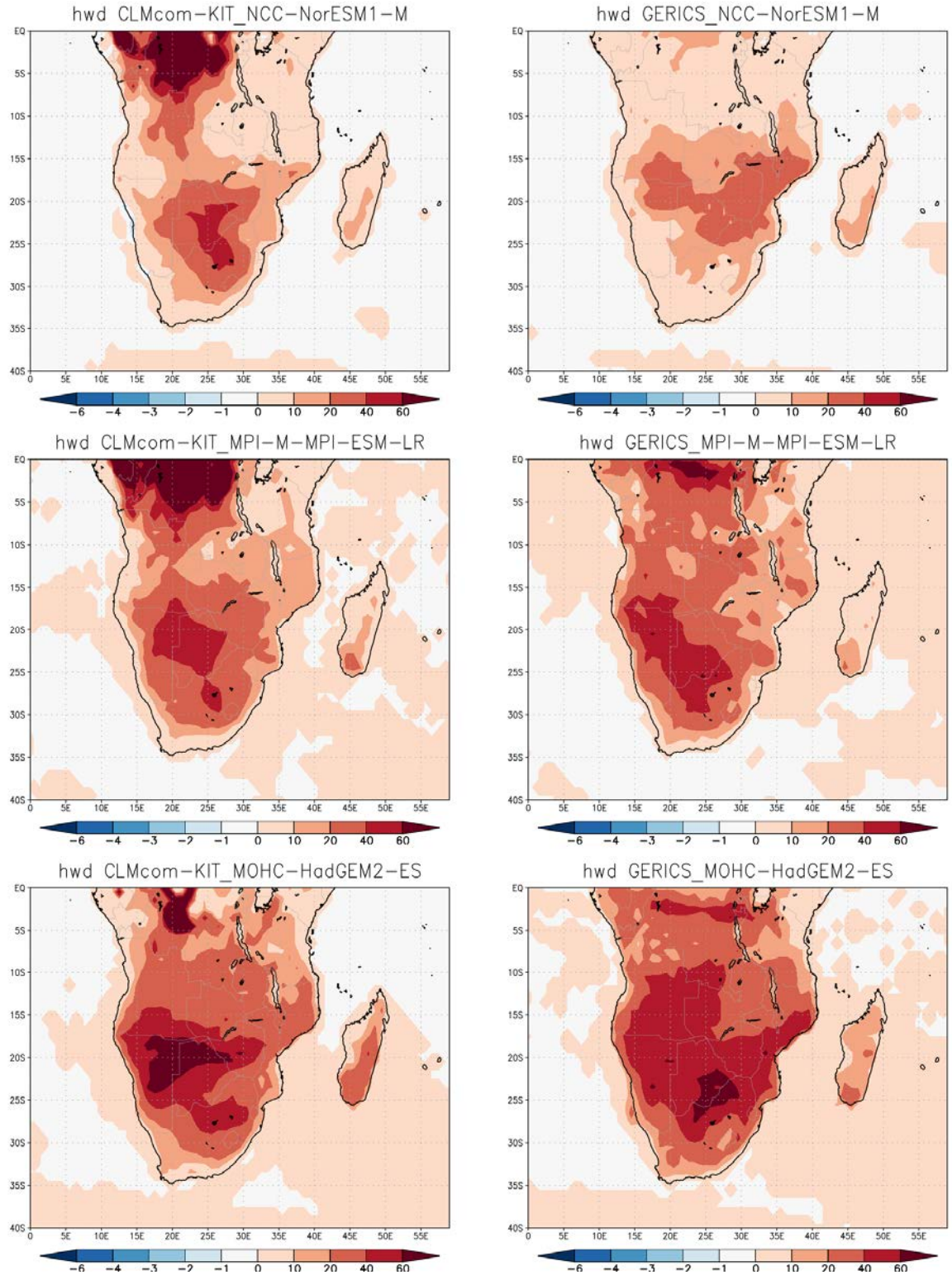


Figure 87: CORDEX-core ensemble projected change in the number of heat-wave days over southern Africa at 8 km resolution, for the time-slab 2080-2099 relative to 2006-2015, under low mitigation.

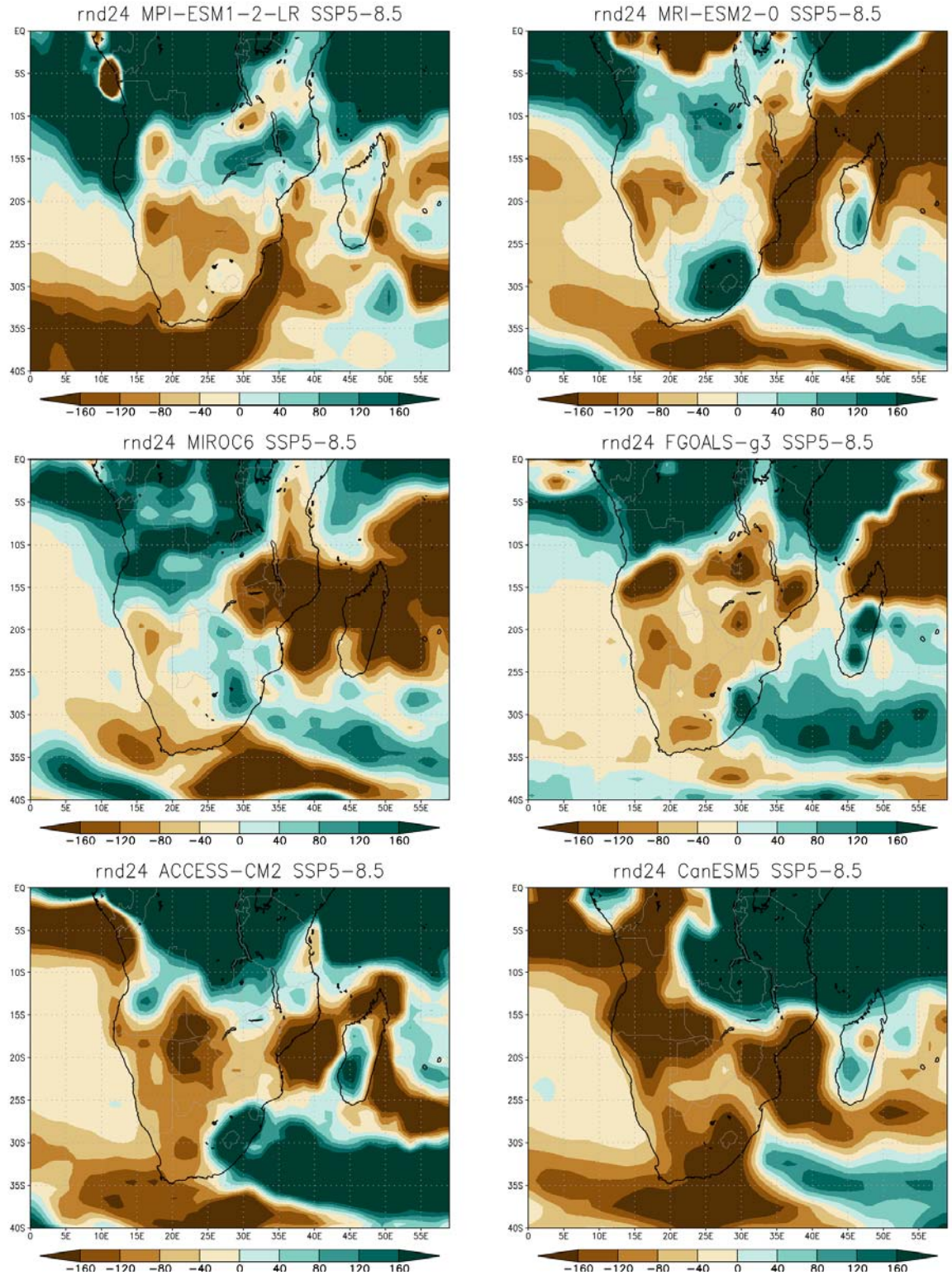


Figure 88: CMIP6 ensemble projected change in annual rainfall (mm) over southern Africa, for the time-slab 2080-2099 relative to 1961-1980, under low mitigation.

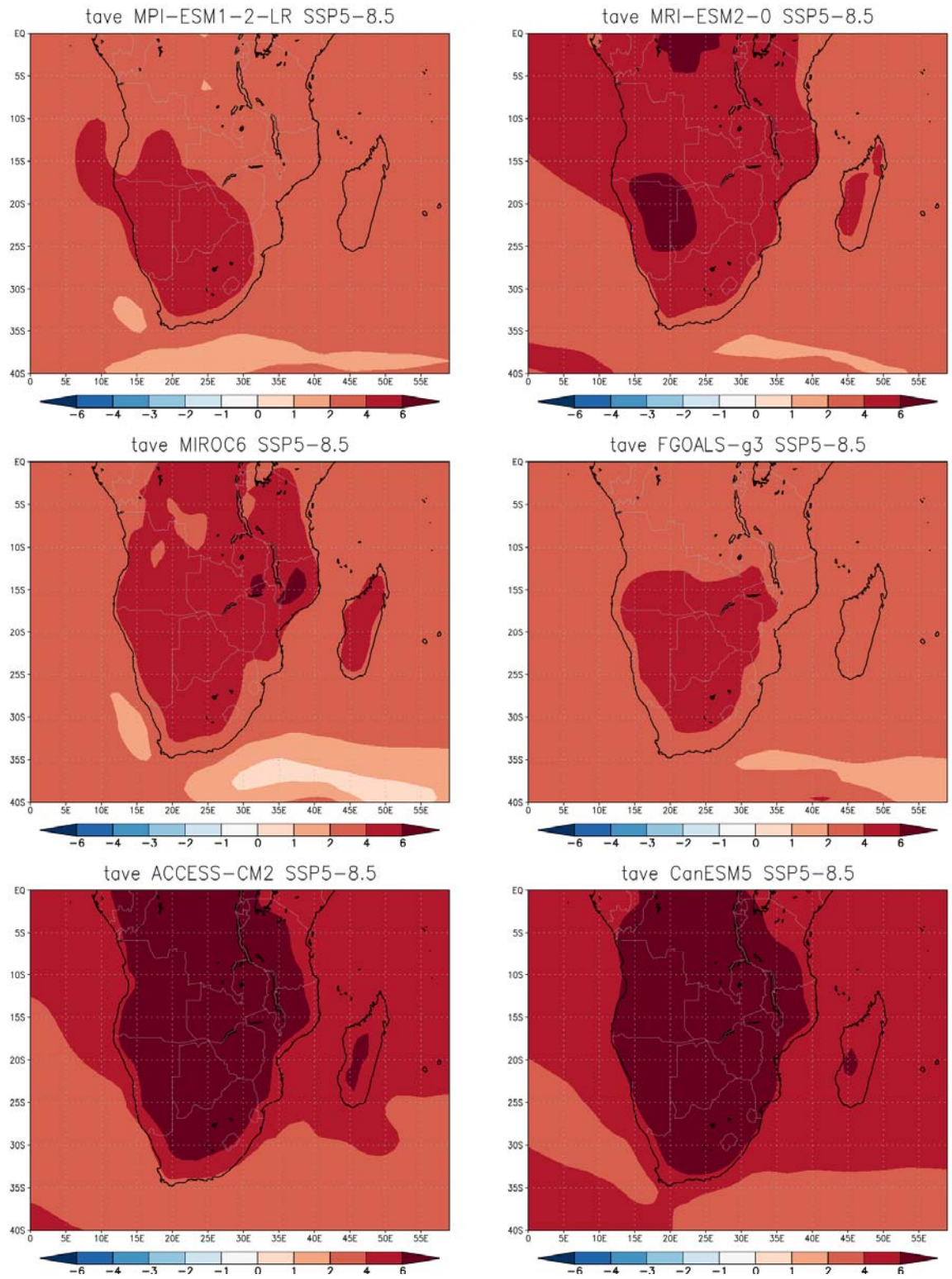


Figure 89: CMIP6 ensemble projected change in the annual average temperature (°C) over southern Africa, for the time-slab 2080-2099 relative to 1961-1980, under low mitigation.

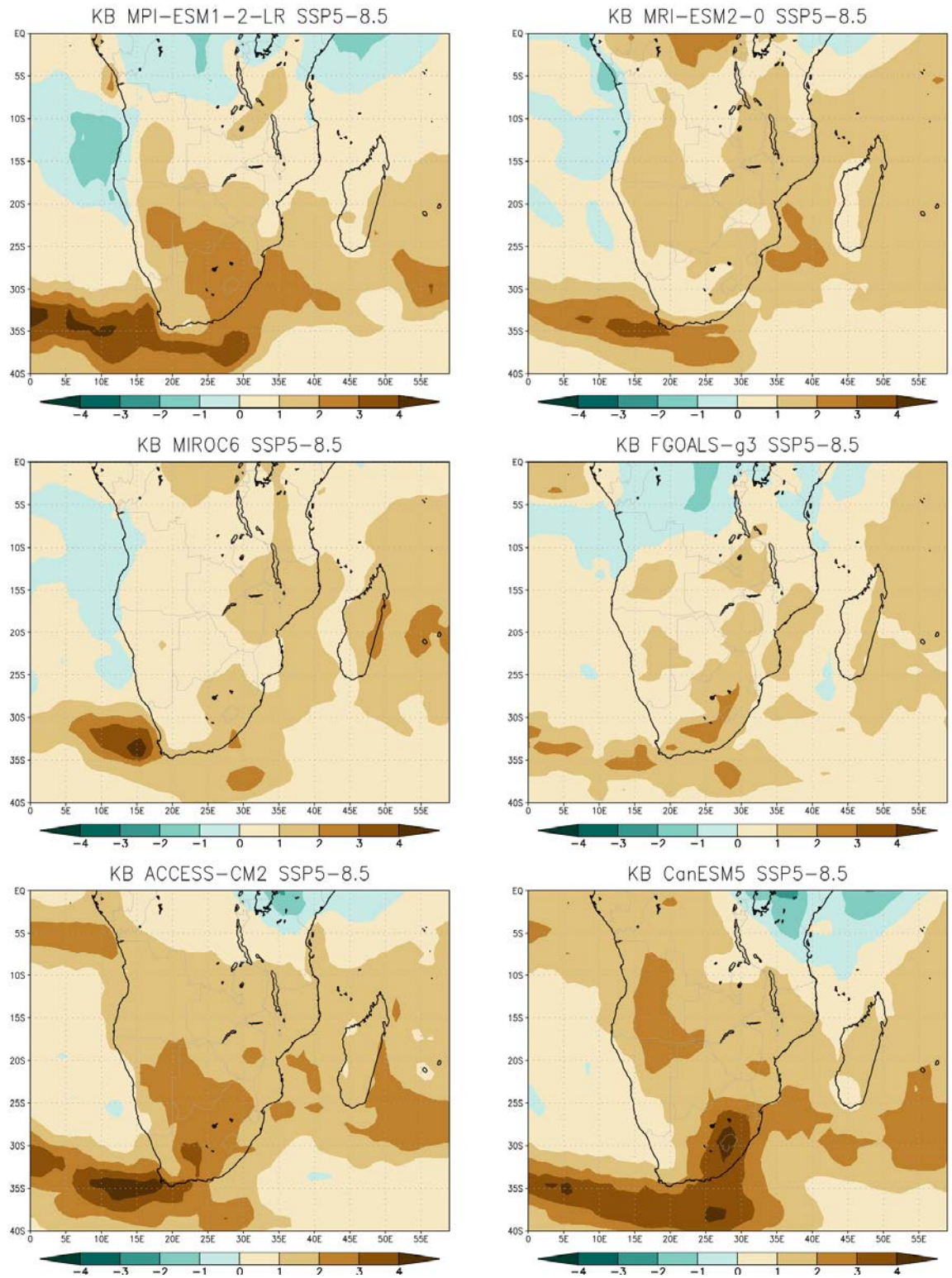


Figure 90: CMIP6 ensemble projected change in the Keetch-Byram drought index over southern Africa, for the time-slab 2080-2099 relative to 1961-1980, under low mitigation.

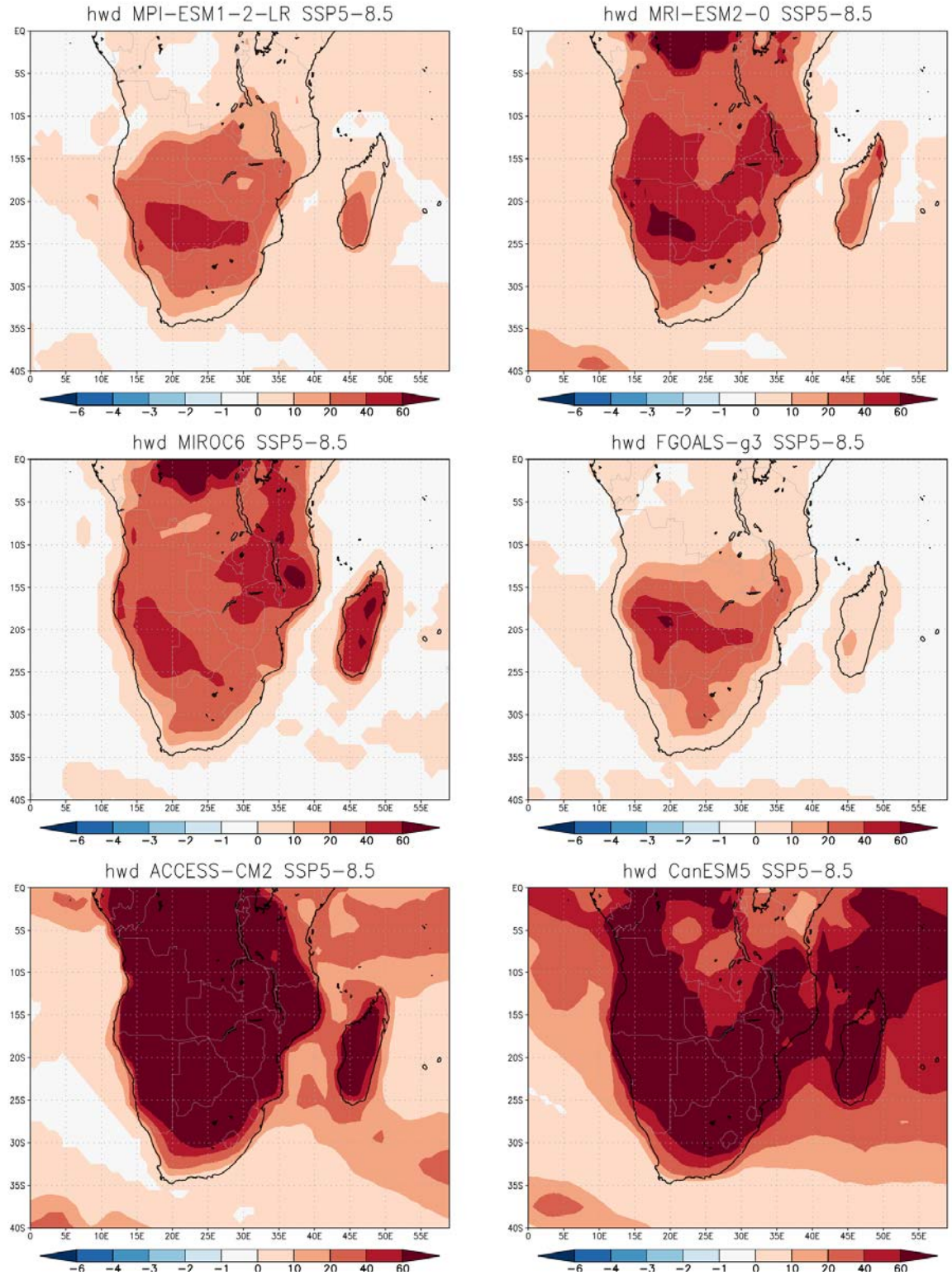


Figure 91: CMIP6 ensemble projected change in the number of heat-wave days over southern Africa at 8 km resolution, for the time-slab 2080-2099 relative to 1961-1980, under low mitigation.

5 Summary and conclusions

This deliverable reports the quality assessment and post-processing work performed on the seasonal, decadal and climate projection data sets in tasks 4.1, 'Seasonal forecast quality assessment' and 4.2, 'Climate projections and decadal assessment'. It covers the WP4 main objectives: (i) improvement of the understanding of the seasonal predictability of the Essential Climate Variables, ECV (ii) improvement of the forecast performance through the application of bias correction approaches and, (iii) assessment of the long-term influences of climate in the areas of the case studies.

The verification of Essential Climate Variables (mean, maximum and minimum temperature as well as precipitation) both for seasonal and decadal climate predictions has shown that the performance in the SADC region is highly dependent on the forecast system, variable, season, area of interest, verification metric and bias correction applied. This finding highlights the potential advantage of providing users with predictions from multiple forecasting systems because it enables them to optimise their choices considering the specific temporal and spatial performance for each individual context. Regarding the bias correction strategies, they should be selected depending on the specific forecast aspect that needs improvement (which strongly correlates to users' decision-making framework). Thus, the results show there is no calibration method that simultaneously improves all the aspects of forecast quality, as each approach is focused on improving the quality measured by a specific metric. Consequently, a systematic forecast quality assessment is always needed to identify the best approach for calibrating the raw simulations for each use-case. With the aim to ease the access to this information to the Focus-Africa users, all the information generated in this systematic assessment is available in a R shiny app (https://earth.bsc.es/shiny/FOCUS-Africa/).

Concerning the study of seasonal extremes, it has been conducted by proposing a new bias correction method based on extending the classic quantile mapping (QM) by improving the distribution tail with a generalized extreme value distribution (GEV). This approach has been applied to temperature and precipitation predictions from three seasonal forecasts systems and the outcomes have been compared with both raw seasonal forecasts and forecasts corrected with the classic QM. To specifically assess the impact on extremes, the focus has been on two particular percentiles, 97th and 99th. In general, the impact of the proposed bias correction technique has been positive, as it has produced improvements even compared to the classic QM. This has been ascertained by looking at the bias, i.e., the difference between the predicted and the reference data (ERA5), which decreased after the correction. Furthermore, RMSE also decreased by up to two orders of magnitude for both variables. For temperature, the effect is similar in both bias correction approaches (classic QM and new), whereas for precipitation, the proposed bias correction produces two different effects: the positive one is the lowering of the error, both in terms of bias and RMSE; the negative, is that it decreases the percentage of ensemble members in agreement with the extreme events detected by ERA5. This contrast indicates an ability of the bias-corrected seasonal forecasts to accurately detect extremes, but a limited skill in placing them on the timeline. Since this also happens with the classic QM, it can be concluded that these techniques allow better identification of the extreme precipitation events, but not of their temporal distribution.

Turning to the revision of the seasonal methodologies in SARCOF and SWIOCOF, the current forecast workflows of the SADC NMHSs generally use the Climate Forecasting Tool (CFT). The CFT uses both

linear regression and multi-linear regression approaches (with Artificial Neural Network, ANN). This is a great improvement compared to previous strategies (the Climate Predictability Tool, still in use) because it allows experts to try out different predictors for temperature and rainfall and, additionally, it is also able to issue both regional and station forecasts with their corresponding skill. However, experts are still advised to run their own statistical models before looking at what the global modelling centres are giving in terms of precipitation forecasts (i.e. to check their coherence). One limitation of these workflows is the scarcity of available documentation and monitoring of drivers of climate variability, their predictability, use of teleconnections and known climate interactions within and outside SADC. For example, interpretation of the outputs of operational global forecasting systems are limited due to unavailable or inaccessible documentation on models' products intercomparison and comparison with observations driving the southern Africa climate variability and change. Another aspect that is still not fully implemented is the verification of the SARCOF forecasts in accordance with standard and user relevant schemes. In fact, user driven verification should facilitate the understanding, interpretation and use of seasonal forecasts in climate sensitive sectors and, hence, their adoption. That said, the NMHSs mainly use visual methods for verifying their outlooks (although they are open to be trained in other available quantitative forecast assessment methods like RPSS).

Since SARCOF forecasts are widely used as a source of seasonal climate information, there has been a specific study analysing the seasonal forecasts for river basins in the SADC region by comparing SARCOF with dynamical forecasts from the Copernicus Climate Data Store. The results have shown that the SARCOF forecasts have very poor reliability as they are strongly biased towards forecasting the normal tercile. They do show skill for tercile categories of above normal and below normal, but it has been argued that this skill is a spurious effect arising from a combination of fixed tercile probabilities and forecast bias. Dynamical forecasts, on the other hand, have shown low levels of skill in general, apart from basins in the coastal region of Tanzania and to the south of 15deg South. That skill is manifested by individual forecasting systems - namely SEAS5, GEM5-NEMO in OND and SEAS5, CFSv2, SPS3.5 and GCFSv2.1, although location of regions differs depending on the forecasting system. As it has been shown in the general analysis of the ECVs, the skill results are highly dependent on the basin, variable, lead time and season and, thus, an interactive tool allowing for exploration of forecasts from the analysed systems and their skill is being developed, and its draft version is accessible through https://cip.csag.uct.ac.za/forecast/sadc-basin.html.

Progressing towards climate projections, the substantial increases in temperature for southern Africa across the CCAM, CORDEX-core, CMIP5 and CMIP6 ensembles, in combination with general reductions in rainfall, imply that there will be great challenges for future regional adaptation. The analysis conducted in this deliverable along with the one carried out in D3.1, allow to propose some recommendations: (i) the CORDEX models should be used for climate projections over the case study countries (i.e. by using as many of the ensemble members as possible to account for the widest range of uncertainties) (ii) when GCM-RCM combinations are duplicated across resolutions (e.g., CCCMA-CANRCM4 and the updated REMO and CCLM RCMs), only use the higher resolution members (iii) for precipitation - due to the larger dependence on the RCM over the driving model - it is possible to take an ensemble average of all the GCM-driven simulations for each RCM, resulting in a single representation of each regional model.

Therefore, one of the main risks the region faces is a more frequent occurrence of multi-year droughts, in combination with increases in very hot days, heat-wave days and high fire-danger days. In fact, the long-term (2080-2099) increases in temperature and extreme temperature events under low mitigation are likely to have wide-spread and substantial impacts in southern and eastern Africa (i.e. for maize crops and cattle industry). Actually, the climate change projections analysed in this report are indicative that such multi-year droughts and higher temperatures are likely to occur more frequently in the near- and mid-term, becoming a dominant feature of southern African climate in the long-term. Another feature is that changes in rainfall totals and extreme rainfall events are likely to occur over the eastern escarpment, consistently from the near-term through to the long-term. A feature of note is that some of the projections are indicative of an increase in total rainfall and extreme events over southern Mozambique, already in the near-term. These changes may be attributed to the more frequent landfall of tropical lows and cyclones in this projection subset. Category 4 and 5 hurricanes have only been detected in the last 15 years in the southwest Indian Ocean. Thus, considering this information, the future landfall of an intense tropical cyclone at Maputo, inland movement of such a system into the Limpopo river basin, or further southwards over Mpumalanga or northeastern KwaZulu-Natal, becomes a feasible event. Such a situation may be regarded to be of low-probability but high-impact, and should be carefully considered in terms of adaptation and disaster management planning.

Regarding the CMIP6 assessment of agroclimatic indicators related to food security (cereals and legume cropping), the overall outcome is that there is reasonable skill in simulating the seasonal climatology for all the indicators. Yet, significant uncertainties still exist, which needs further future exploration (the performance varied significantly across individual models). The performance of CMIP6 models, though, was reduced when simulating the interannual variability of the seasonal cycle as compared with the spatial pattern of seasonal climatology. It is worth noting that the documented biases and relative errors in this study may not be simply alleviated by increasing the model's horizontal resolution because, for some indicators, moderate resolution models exhibit better performance than higher resolution ones (Akinsanola et al, 2020, 2021). However, given the knowledge of such biases, the CMIP6 climate models can be suitably applied to project future state and able to assist policy makers in their decisions on food security, climate change adaptation and mitigation action (i.e. for particular crop production impact studies).

Finally the specific case-study analysis of Lake Malawi showed that the SOM has been able to characterize the moisture flux states over the region and identify easterly and westerly flows into the catchment area (suggesting that SOMs could be used in a GCM selection process). More specifically, the majority of SOM states are associated with easterly flows, with only 2 nodes showing westerly flows. Nevertheless, large-scale rainfall is associated with westerly flow states significantly contributing to large-scale rainfall over the Lake Malawi catchment as well as to local maxima. Easterly flows, on the other hand, are associated with lower rainfall amounts during the shoulder seasons and summer. In that sense, most GCMs replicate the monthly nodal mapping adequately, although several GCMs did not replicate mappings to the important rain-bearing nodes and were excluded from further analysis. As for the projected monthly mappings of GCM data, they are very similar to those of the historical period for the important rain-bearing nodes indicating that the frequency of synoptic states associated with rainfall will remain consistent in the medium-term future. This no-change signal is

likely a function of the medium-term time horizon selected for the study where the level of global warming is projected to be approximately 1.5 degrees by 2050 (under this scenario).

A detailed analysis of this behaviour shows that the IOD and westerly synoptic flow are drivers of regional rainfall over the Lake Malawi catchment. Linking these atmospheric processes to rainfall in the catchment area has provided a physical explanation of the observed and projected characteristics in rainfall over the study region, where most of the global climate models (GCMs) adequately reproduce these drivers. More specifically, they are able to capture the spatial rainfall anomalies associated with the IOD and ENSO, and that higher rainfall is associated with the positive phase of the IOD and westerly synoptic flow. Although rainfall projections for the catchment associated with a +IOD phase decrease over the period 2030-2059 under the midrange SSP245 scenario, rare and heavy rainfall events are projected to increase. However, at the synoptic scale no large changes in the frequency of occurrence of rain-bearing synoptics are projected. It is important to highlight that these results are particular to the Lake Malawi catchment basin (the teleconnective and synoptic forcings presented in this analysis could have different local and regional rainfall expressions in other parts of the southern or eastern African domains).

The knowledge presented in these deliverable will be used in future tasks 4.3 (‘Implementation of multi-model and downscaling for seasonal forecasts’) and 4.4 (‘Implementation of multi-model and downscaling for climate projections and decadal predictions’) to further improve and tailor the different climate predictions. For instance, the multi-model combination (which consists of merging the predictions provided by several forecast systems) is expected to enhance the quality and reliability of the predictions due to the error compensation and the signal addition that each system sums to the multi-model ensemble. Also, downscaling techniques will improve the information provided to users, as many applications need predictions at regional to local scales.

6 Bibliography

Abba Omar, S., & Abiodun, B. J. (2017). How well do CORDEX models simulate extreme rainfall events over the East Coast of South Africa? Theoretical and Applied Climatology, 128(1–2), 453–464. https://doi.org/10.1007/s00704-015-1714-5

Abiodun, B. J., Mogebisa, T. O., Petja, B., Abatan, A. A., & Roland, T. R. (2020). Potential impacts of specific global warming levels on extreme rainfall events over southern Africa in CORDEX and NEX-GDDP ensembles. International Journal of Climatology, 40(6), 3118–3141. https://doi.org/10.1002/joc.6386

ADB (2019). Southern Africa Economic Outlook 2019. Macroeconomic performance and prospects, African Development Bank , ISBN 978-9938-882-98-8. https://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/2019AEO/REO_2019_-_Southern_africa.pdf

AgERA5, Copernicus Climate Change Service (C3S), Fifth generation of ECMWF atmospheric reanalysis of the global climate for agriculture and ago-ecological studies (2021). Copernicus Climate Change Service Climate Data Store (CDS), July-2021.

Akinsanola, A. A., Kooperman, G. J., Pendergrass, A. G., Hannah, W. M., and Reed, K. A. (2020). Seasonal representation of extreme precipitation indices over the United States in CMIP6 present-day simulations. United States: N. p., 2020. Web. doi:10.1088/1748-9326/ab92c1.

Akinsanola, Akintomide & Ongoma, Victor & Kooperman, Gabriel. (2021). Evaluation of CMIP6 models in simulating the statistics of extreme precipitation over Eastern Africa. Atmospheric Research. 254. 105509. 10.1016/j.atmosres.2021.105509.

Allan, R., and Ansell, T. (2006): A New Globally Complete Monthly Historical Gridded Mean Sea Level Pressure Dataset (HadSLP2): 1850-2004. Journal of Climate, 19, 5816-5842. https://doi.org/10.1175/JCLI3937.1

Allan, R. P., and coauthors (2020). Advances in understanding large-scale responses of the water cycle to climate change. Ann. N. Y. Acad. Sci., 1472, 49–75, https://doi.org/10.1111/nyas.14337.

Almazroui, M., Ashfaq, M., Islam, M. N., Rashid, I. U., Kamil, S., Abid, M. A., … Sylla, M. B. (2021). Assessment of CMIP6 Performance and Projected Temperature and Precipitation Changes Over South America. Earth Systems and Environment, 5(2), 155–183. https://doi.org/10.1007/s41748-021-00233-6

Almazroui, M., Saeed, F., Saeed, S., Ismail, M., Ehsan, M. A., Islam, M. N., … Nadeem, I. (2021). Projected Changes in Climate Extremes Using CMIP6 Simulations Over SREX Regions. Earth Systems and Environment, 5(3), 481–497. https://doi.org/10.1007/s41748-021-00250-5

Archer, E., Engelbrecht, F., Hänsler, A., Landman, W., Tadross, M., & Helmschrot, J. (2018). Seasonal prediction and regional climate projections for southern Africa. Biodiversity & Ecology, 6, 14–21. https://doi.org/10.7809/b-e.00296

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., et al. (2021). NorCPM1 and its contribution to CMIP6 DCPP, Geoscientific Model Development. https://doi.org/10.5194/gmd-14-7073-2021

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P. A., et al (2021). Assessment of a full-field initialised decadal climate prediction system with the CMIP6 version of EC-Earth. Earth System Dynamics. https://doi.org/10.5194/esd-12-173-2021

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6, Geosci. Model Dev., 9, 3751–3777, https://doi.org/10.5194/gmd-9-3751-2016

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., et al (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. Journal of Advances in Modeling Earth Systems. https://doi.org/10.1029/2019MS002010

Cannon, A. J., Piani, C., & Sippel, S. (2020). Bias correction of climate model output for impact models. In *Climate Extremes and Their Implications for Impact and Risk Assessment* (pp. 77-104). Elsevier.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., et al. (2018). Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model. Journal of Advances in Modeling Earth Systems. https://doi.org/10.1029/2018MS001369

Cottrill, A., Hendon, H. H., Lim, E. P., Langford, S., Shelton, K., Charles, A., McClymont, D., Jones, D., & Kuleshov, Y. (2013). Seasonal Forecasting in the Pacific Using the Coupled Model POAMA-2. Weather and Forecasting, 28(3), 668–680. https://doi.org/10.1175/WAF-D-12-00072.1

CSIRO and Bureau of Meteorology (2015). Climate Change in Australia. Information for Australia's Natural Resource Management Regions: Technical Report. CSIRO and Bureau of Meteorology, Australia

Diallo, I., Giorgi, F., Sukumaran, S., Stordal, F., & Giuliani, G. (2015). Evaluation of RegCM4 driven by CAM4 over Southern Africa: mean climatology, interannual variability and daily extremes of wet season temperature and precipitation. Theoretical and Applied Climatology, 121(3–4), 749–766. https://doi.org/10.1007/s00704-014-1260-6

Dedekind Z., Engelbrecht F.A. and Van der Merwe J. (2016). Model simulations of rainfall over southern Africa and its eastern escarpment. Water SA 42 129-143. http://dx.doi.org/10.4314/wsa.v42i1.13.

DelSole, T., & Tippett, M. K. (2016). Forecast Comparison Based on Random Walks, Monthly Weather Review, 144(2), 615-626. https://doi.org/10.1175/MWR-D-15-0218.1

Doblas-Reyes, F. J., Hagedorn, R., and Palmer, T. N. (2005) 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, https://doi.org/10.3402/tellusa.v57i3.14658

Doblas-Reyes, F. J., Andreu-Burillo, I., Chikamoto, Y. et al. Initialized near-term regional climate change prediction. Nat Commun 4, 1715 (2013). https://doi.org/10.1038/ncomms2704

Doblas-Reyes, F. J. *et al.* (2013) 'Seasonal climate predictability and forecasting: Status and prospects', *Wiley Interdisciplinary Reviews: Climate Change*, 4(4), pp. 245–268. doi: 10.1002/wcc.217.

D'Onofrio, D., Palazzi, E., von Hardenberg, J., Provenzale, A., & Calmanti, S. (2014). Stochastic rainfall downscaling of climate models. *Journal of Hydrometeorology*, *15*(2), 830-843.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J.-P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégoz, M., Miller, P. A., Moreno-Chamarro, E., Nieradzik, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wårlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q. (2022). The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, Geosci. Model Dev., 15, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022, 2022.

Dosio, A., Jones, R. G., Jack, C., Lennard, C., Nikulin, G., & Hewitson, B. (2019). What can we know about future precipitation in Africa? Robustness, significance and added value of projections from a large ensemble of regional climate models. Climate Dynamics, 53(9–10), 5833–5858. https://doi.org/10.1007/s00382-019-04900-3

Dosio A., Jury M.W., Almazroui M., Ashfaq M., Diallo I., Engelbrecht F.A., Klutse N.A.B., Lennard C., Pinto I., Sylla M.B. and Tamoffo A.T. (2021). Projected future daily characteristics of African precipitation based on global (CMIP5, CMIP6) and regional (CORDEX, CORDEX-CORE) climate models. Climate Dynamics 57 3135–3158. https://doi.org/10.1007/s00382-021-05859-w.

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., et al. (2020). The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. Journal of Advances in Modeling Earth Systems, 12, e2019MS002015. https://doi.org/10.1029/2019MS002015

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

Eccel, E., A. L. Zollo, P. Mercogliano, and R. Zorer (2016). Simulations of quantitative shift in bio-climatic indices in the viticultural areas of Trentino (Italian Alps) by an open source R package. Comput. Electron. Agric., 127, 92–100, https://doi.org/10.1016/J.COMPAG.2016.05.019.

Edward M. Mugalavai, Emmanuel C. Kipkorir, Dirk Raes, Manchiraju S. Rao (2008). Analysis of rainfall onset, cessation and length of growing season for western Kenya. Agricultural and Forest Meteorology, Volume 148, Issues 6–7, 2008, Pages 1123-1135, ISSN 0168-1923, https://doi.org/10.1016/j.agrformet.2008.02.013.

Engelbrecht, F. A., Marean, C. W., Cowling, R. M., Engelbrecht, C. J., Neumann, F. H., Scott, L., … Difford, M. (2019). Downscaling Last Glacial Maximum climate over southern Africa. Quaternary Science Reviews, 226, 105879. https://doi.org/10.1016/j.quascirev.2019.105879

Engelbrecht F.A., Adegoke J., Bopape M-J., Naidoo M., Garland R., Thatcher M., McGregor J., Katzfey J., Werner M., Ichoku C. and Gatebe C. (2015). Projections of rapidly rising surface temperatures over Africa under low mitigation. Env. Res. Letters. 10 085004. https://doi.org/10.1088/1748-9326/10/8/085004.

Evans, J. P., Ji, F. Abramowitz, G. and Ekstrom, M. (2013). Optimally choosing small ensemble members to produce robust climate simulations. Environmental Research Letters, 8(4): 044050.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development, 9(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

FAO (1978). Forestry for Local Community Development Food and Agriculture Organization of the United Nation (FAO), FAO Forestry paper, No 7, Rome.

FAO (1977). Crop water requirements. FAO Irrigation and Drainage Paper No. 24, by Doorenbos J and W.O. Pruitt. FAO, Rome, Italy.

FAO (1986). Early Agrometeorological crop yield forecasting. FAO Plant Production and Protection paper No. 73, by M. Frère and G.F. Popov. FAO, Rome, Italy.

Ferro, C.A.T. (2014), Fair scores for ensemble forecasts. Q.J.R. Meteorol. Soc., 140: 1917-1923. https://doi.org/10.1002/qj.2270

Finney, D. L., Marsham, J. H., Jackson, L. S., Kendon, E. J., Rowell, D. P., Boorman, P. M., … Senior, C. A. (2019). Implications of Improved Representation of Convection for the East Africa Water Budget Using

a Convection-Permitting Model. Journal of Climate, 32(7), 2109–2129. https://doi.org/10.1175/JCLI-D-18-0387.1

Fitchett J.M. (2018). Recent emergence of CAT5 tropical cyclones in the South Indian Ocean. South African Journal of Science 114 11-12. http://dx.doi.org/10.17159/sajs.2018/4426.

Fortin, V., Abaza, M., Anctil, F., & Turcotte, R. (2014). Why Should Ensemble Spread Match the RMSE of the Ensemble Mean?, Journal of Hydrometeorology, 15(4), 1708-1713. https://doi.org/10.1175/JHM-D-14-0008.1

Funk, C., L. Harrison, L. Alexander, P. Peterson, A. Behrangi, and G. Husak (2019). Exploring trends in wet-season precipitation and drought indices in wet, humid and dry regions. Environ. Res. Lett., 14, 115002, https://doi.org/10.1088/1748-9326/ab4a6c.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific Data, 2, 150066. https://doi.org/10.1038/sdata.2015.66

Garland R.M., Matooane M., Engelbrecht F.A., Bopape M-J.M., Landman W.A., Naidoo M., Van der Merwe J. And Wright C.Y. (2015). Regional projections of extreme apparent temperature days in Africa and the related potential risk to human health. International Journal of Environmental Research and Public Health 10/2015 12 12577-12604. https://doi.org/10.3390/ijerph121012577.
Gnitou G.T., Tan G., Niu R. and Nooni I.K. (2021). Assessing past climate biases and the added value of cordex-core precipitation simulations over africa. Remote Sens 13 2058. https:// doi. org/ 10. 3390/ rs131 12058.

Giorgi, F., F. Raffaele, and E. Coppola (2019). The response of precipitation characteristics to global warming from climate projections. Earth Syst. Dyn., 10, 73–89, https://doi.org/10.5194/esd-10-73-2019.

Gitz, Vincent & Meybeck, Alexandre & Lipper, Leslie & Young, Cassandra & Braatz, Susan. (2016). Climate change and food security: risks and responses.

Gnitou, G. T., Tan, G., Niu, R., & Nooni, I. K. (2021). Assessing Past Climate Biases and the Added Value of CORDEX-CORE Precipitation Simulations over Africa. Remote Sensing, 13(11), 2058. https://doi.org/10.3390/rs13112058

Goddard, L., Kumar, A., Solomon, A. et al. (2013). A verification framework for interannual-to-decadal predictions experiments. Clim Dyn 40, 245–272. https://doi.org/10.1007/s00382-012-1481-2

Gualdi, S., et al. (2020) The new CMCC operational seasonal prediction system. CMCC Research Paper, (RP0288). https://doi.org/10.25424/CMCC/SPS3.5

Hariadi, M. H., van der Schrier, G., Steeneveld, G.-J., Sopaheluwakan, A., Tank, A. K., Roberts, M. J., Moine, M.-P., Bellucci, A., Senan, R., Tourigny, E., & Putrasahan, D. (2022). Evaluation of onset, cessation and seasonal precipitation of the Southeast Asia rainy season in CMIP5 regional climate models and HighResMIP global climate models. International Journal of Climatology, 42(5), 3007–3024. https://doi.org/10.1002/joc.7404

Harrison, Laura & Michaelsen, Joel & Funk, Chris & Husak, Gregory. (2011). Effects of temperature changes on maize production in Mozambique. Climate Research. 46. 211-222. 10.3354/cr00979.

Hazeleger, W., Guemas, V., Wouters, B., Corti, S., Andreu–Burillo, I., Doblas–Reyes, F. J., Wyser, K., and Caian, M. (2013), Multiyear climate predictions using two initialization strategies, Geophys. Res. Lett., 40, 1794– 1798, https://doi.org/10.1002/grl.50355

Hersbach, H, Bell, B, Berrisford, P, et al. (2020). The ERA5 global reanalysis. Q J R Meteorol Soc., 146: 1999– 2049. https://doi.org/10.1002/qj.3803

Hoegh-Guldberg, O., D. Jacob, M. Taylor, M. Bindi, S. Brown, I. Camilloni, A. Diedhiou, R. Djalante, K.L. Ebi, F. Engelbrecht, J. Guiot, Y. Hijioka, S. Mehrotra, A. Payne, S.I. Seneviratne, A. Thomas, R. Warren, and G. Zhou, 2018: Impacts of 1.5ºC Global Warming on Natural and Human Systems. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)].

Horowitz, H. M., Garland, R. M., Thatcher, M., Landman, W. A., Dedekind, Z., van der Merwe, J., & Engelbrecht, F. A. (2017). Evaluation of climate model aerosol seasonal and spatial variability over Africa using AERONET. Atmospheric Chemistry and Physics, 17(22), 13999–14023. https://doi.org/10.5194/acp-17-13999-2017

IPCC (2019). Summary for Policymakers. In: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems [P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.- O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.)]. In press

IPCC (2021). Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. In Press.

James, R., Washington, R., Abiodun, B., Kay, G., Mutemi, J., Pokam, W., Hart, N., Artan, G., & Senior, C. (2018). Evaluating Climate Models with an African Lens, Bulletin of the American Meteorological Society, 99(2), 313-336. Retrieved Jul 25, 2022, from https://journals.ametsoc.org/view/journals/bams/99/2/bams-d-16-0090.1.xml

Johnson, C., & Bowler, N. (2009). On the Reliability and Calibration of Ensemble Forecasts. Monthly Weather Review, 137(5), 1717–1720. https://doi.org/10.1175/2009MWR2715.1

Johnson SJ, et al. (2019) SEAS5: the new ECMWF seasonal forecast system. Geosci Model Dev 12(3):1087–1117. https://doi.org/10.5194/gmd-12-1087-2019

Jolliffe, I. T. and Stephenson, D. B. (2012). Forecast verification: a practitioner's guide in atmospheric science. John Wiley and Sons. ISBN: 978-0-470-66071-3

Keetch J.J. and Byram G.M. (1968). A drought index for fire control. Res. Pap. SE-38 (Asheville, NC: US Department of Agriculture, Forest Service, Southeastern Forest Experiment Station) p 32 (revised November 1988).

Kendon, E. J., Stratton, R. A., Tucker, S., Marsham, J. H., Berthou, S., Rowell, D. P., & Senior, C. A. (2019). Enhanced future changes in wet and dry extremes over Africa at convection-permitting scale. Nature Communications, 10(1), 1794. https://doi.org/10.1038/s41467-019-09776-9

Kharin, V. V., & Zwiers, F. W. (2003). On the ROC Score of Probability Forecasts, Journal of Climate, 16(24), 4145-4150. https://doi.org/10.1175/1520-0442(2003)016%3C4145:OTRSOP%3E2.0.CO;2

Kim, S.-K. Min, X. Zhang, J. Sillmann, M. Sandstad (2020). Evaluation of the CMIP6 multi-model ensemble for climate extreme indices Weather Clim. Extremes, 29, Article 100269

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. Journal of the Meteorological Society of Japan. https://doi.org/10.2151/jmsj.2015-001

Kothe, S., Panitz, H.-J., & Ahrens, B. (2014). Analysis of the radiation budget in regional climate simulations with COSMO-CLM for Africa. Meteorologische Zeitschrift, 23(2), 123–141. https://doi.org/10.1127/0941-2948/2014/0527

Kruger A.C. and Nxumaloa M. (2016). Surface temperature trends from homogenized time series in South Africa: 1931–2015. Int. J. Climatol. DOI: 10.1002/joc.4851.

Kruger A.C. and Sekele S.S. Trends in extreme temperature indices in South Africa: 1962–2009. Int. J. Climatol. 33 661–676.

Lee J.Y., J. Marotzke, G. Bala, L. Cao, S. Corti, J.P. Dunne, F. Engelbrecht, E. Fischer, J.C. Fyfe, C. Jones, A. Maycock, J. Mutemi, O. Ndiaye, S. Panickal, T. Zhou (2021). Future Global Climate: Scenario-Based Projections and Near-Term Information. In: Climate Change 2021: The Physical Science Basis.

Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. Maycock, T. Waterfield, 46O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press.

Lledó, Ll, et al. (2019) 'Seasonal forecasts of wind power generation.' Renewable Energy 143: 91-100. https://doi.org/10.1016/j.renene.2019.04.135

Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E., et al. (2022). CMIP6 simulations with the CMCC Earth System Model (CMCC-ESM2). Journal of Advances in Modeling Earth Systems, 14, e2021MS002814. https://doi.org/10.1029/2021MS002814

Luo, Jing-Jia, et al. (2011) 'Impact of global ocean surface warming on seasonal-to-interannual climate prediction.' Journal of Climate 24.6: 1626-1646. https://doi.org/10.1175/2010JCLI3645.1

Magadzire, T. (2012). Evaluation of SARCOF products from 2001 to 2012. [PowerPoint slides]. FEWSNET. SARCOF Mid-Season Correction Update. 11 December 2012.

Maraun, D., & Widmann, M. (2018). Statistical downscaling and bias correction for climate research. Cambridge University Press.
Marcos, Raül, et al. (2017) 'Seasonal predictability of water resources in a Mediterranean freshwater reservoir and assessment of its utility for end-users.' Science of the Total Environment 575: 681-691. https://doi.org/10.1016/j.scitotenv.2016.09.080

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO2. Journal of Advances in Modeling Earth Systems. https://doi.org/10.1029/2018MS001400

McGregor, J. L. (2005). C-CAM geometric aspects and dynamical formulation. CSIRO Atmospheric Research Technical Paper.

McSweeney, C. F., Jones, R. G., Lee, R. W., & Rowell, D. P. (2015). Selecting CMIP5 GCMs for downscaling over multiple regions. Climate Dynamics, 44(11–12), 3237–3260. https://doi.org/10.1007/s00382-014-2418-8

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., and Lawrimore, J. H. (2018). The Global Historical Climatology Network Monthly Temperature Dataset, Version 4. Journal of Climate. https://doi.org/10.1175/JCLI-D-18-0094.1

Merryfield WJ, et al. (2013) The canadian seasonal to interannual prediction system. Part I: models and initialization. Mon Weather Rev 141:2910–2945. https://doi.org/10.1175/MWR-D-12-00216.1

Moise, Aurel & Wilson, Louise & Grose, Michael & Whetton, Penny & Watterson, I.G. & Bhend, Jonas & Bathols, J. & Hanson, L. & Erwin, Tim & Bedin, T. & Heady, C. & Rafter, Tony (2015). Evaluation of

CMIP3 and CMIP5 Models over the Australian Region to Inform Confidence in Projections. Australian Meteorological and Oceanographic Journal. 65. 19-53. 10.22499/2.6501.004.

Moriasi, D.N., Arnold, J.G., Liew, M W Van, Bingner, R.L., Harmel, R.D., Veith, T.L., Van Liew, Michael W, Bingner, R.L., Harmel, R.D., Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50, 885–900.

Müller, W., Jungclaus, J., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., et al. (2018). A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM 1.2-HR). Journal of Advances in Modeling Earth Systems. https://doi.org/10.1029/2017MS001217

Nkrumah, F.; Quagraine, K.A.; Quagraine, K.T.; Wainwright, C.; Quenum, G.M.L.D.; Amankwah, A.; Klutse, N.A.B. (2022). Performance of CMIP6 HighResMIP on the Representation of Onset and Cessation of Seasonal Rainfall in Southern West Africa. Atmosphere, 13, 999. https://doi.org/10.3390/atmos13070999

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., … Solecki, W. (2017). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. Global Environmental Change, 42, 169–180. https://doi.org/10.1016/j.gloenvcha.2015.01.004

Otto F.E.L., Zachariah M., Wolski P., Pinto I., Barimalala R., Nhamtumbo B., Bonnet R., Vautard R., Philip S., Kew S., Luu L.N., Heinrich D., Vahlberg M., Singh R., Arrighi J., Thalheimer L., Van Aalst M., Li S., Sun J., Vecchi G. and Harrington L.J. (2022). Climate change increased rainfall associated with tropical cyclones hitting highly vulnerable communities in Madagascar, Mozambique & Malawi. Word Weather Attribution Service. https://www.worldweatherattribution.org/climate-change-increased-rainfall-associated-with-tropical-cyclones-hitting-highly-vulnerable-communities-in-madagascar-mozambique-malawi/

Panitz, H.-J., Dosio, A., Büchner, M., Lüthi, D., & Keuler, K. (2014). COSMO-CLM (CCLM) climate simulations over CORDEX-Africa domain: analysis of the ERA-Interim driven simulations at 0.44° and 0.22° resolution. Climate Dynamics, 42(11–12), 3015–3038. https://doi.org/10.1007/s00382-013-1834-5

Pérez-Zanón, N., Caron, L.-P., Terzago, S., Van Schaeybroeck, B., Lledó, L., Manubens, N., Roulin, E., Alvarez-Castro, M. C., Batté, L., Delgado-Torres, C., Domínguez, M., von Hardenberg, J., Sánchez-García, E., Torralba, V., and Verfaillie, D. (2022). The CSTools Toolbox: from Climate Forecasts to Climate Forecast Information. Geoscientific Model Development. https://doi.org/10.5194/gmd-2021-368

Piani, C., Haerter, J. O., & Coppola, E. (2010). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology* 99, 187-192

Pinto, I., Lennard, C., Tadross, M., Hewitson, B., Dosio, A., Nikulin, G., … Shongwe, M. E. (2016). Evaluation and projections of extreme precipitation over southern Africa from two CORDEX models. Climatic Change, 135(3–4), 655–668. https://doi.org/10.1007/s10584-015-1573-1

Prein, A. F., Gobiet, A., Truhetz, H., Keuler, K., Goergen, K., Teichmann, C., … Jacob, D. (2016). Precipitation in the EURO-CORDEX $$0.11^{\circ }$$ 0 . 11 ∘ and $$0.44^{\circ }$$ 0 . 44 ∘ simulations: high resolution, high benefits? Climate Dynamics, 46(1–2), 383–412. https://doi.org/10.1007/s00382-015-2589-y

Ranasinghe, R., A.C. Ruane, R. Vautard, N. Arnell, E. Coppola, F.A. Cruz, S. Dessai, A.S. Islam, M. Rahimi, D. Ruiz Carrascal, J. Sillmann, M.B. Sylla, C. Tebaldi, W. Wang, and R. Zaaboul, 2021: Climate Change Information for Regional Impact and for Risk Assessment. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1767–1926, doi:10.1017/9781009157896.014.

Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., et al. (2007). Climate models and their evaluation. In Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the IPCC (FAR) pp. 589–662. Cambridge University Press.

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., … Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change, 42, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Rockström, J., and Coauthors (2010). Managing water in rainfed agriculture—The need for a paradigm shift. Agric. Water Manage., 97, 543–550, https://doi.org/10.1016/j.agwat.2009.09.009.

Romain Marteau, Benjamin Sultan, Vincent Moron, Agali Alhassane, Christian Baron, Seydou B. Traoré (2011). The onset of the rainy season and farmers' sowing strategy for pearl millet cultivation in Southwest Niger, Agricultural and Forest Meteorology, Volume 151, Issue 10, 1356-1369, ISSN 0168-1923, https://doi.org/10.1016/j.agrformet.2011.05.018.

Roulin, E., & Vannitsem, S. (2015). Post-processing of medium-range probabilistic hydrological forecasting: impact of forcing, initial conditions and model errors. Hydrological Processes, 29(6), 1434–1449. https://doi.org/10.1002/HYP.10259

Saha S, et al. (2014). The NCEP climate forecast system version 2. J Clim 27(6):2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Scaife, A. A. and Smith, D. (2018). A signal-to-noise paradox in climate science. npj Clim Atmos Sci 1, 28. https://doi.org/10.1038/s41612-018-0038-4

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., and Ziese, M. (2018): GPCC Full Data Monthly Product Version 2018 at 1.0°: Monthly Land-Surface Precipitation from Rain-Gauges built on

GTS-based and Historical Data. Global Precipitation Climatology Centre (GPCC, http://gpcc.dwd.de/) at Deutscher Wetterdienst. http://dx.doi.org/10.5676/DWD_GPCC/FD_M_V2018_050

Sellar, A. A., Walton, J., Jones, C. G., Wood, R., Luke Abraham, N., Andrejczuk, M., et al. (2020). Implementation of U.K. Earth System Models for CMIP6. Journal of Advances in Modeling Earth Systems. https://doi.org/10.1029/2019MS001946

Seneviratne, S. I., & Hauser, M. (2020). Regional Climate Sensitivity of Climate Extremes in CMIP6 Versus CMIP5 Multimodel Ensembles. Earth's Future, 8(9). https://doi.org/10.1029/2019EF001474
Senior, C., Finney, D., Owiti, Z., Rowell, D., Marsham, J., Jackson, L., … Misiani, H. (2020). Technical guidelines for using CP4-Africa simulation data, (November). Retrieved from https://zenodo.org/record/4316467

Seneviratne, S.I., X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S.M. Vicente-Serrano, M. Wehner, and B. Zhou (2021). Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1513–1766, doi:10.1017/9781009157896.013.

Shah, H., Hellegers, P., and Siderius, C. (2021). Climate risk to agriculture: a synthesis to define different types of critical moments. Clim. Risk Manage. 34:100378. doi: 10.1016/J.CRM.2021.100378

Shukla S, Husak G, Turner W, Davenport F, Funk C, Harrison L, et al. (2021) A slow rainy season onset is a reliable harbinger of drought in most food insecure regions in Sub-Saharan Africa. PLoS ONE 16(1): e0242883. https://doi.org/10.1371/journal.pone.0242883

Smith, D.M., Eade, R., Scaife, A.A. et al. (2019). Robust skill of decadal climate predictions. npj Clim Atmos Sci 2, 13. https://doi.org/10.1038/s41612-019-0071-y

Sørland, S. L., Brogli, R., Pothapakula, P. K., Russo, E., Van de Walle, J., Ahrens, B., … Thiery, W. (2021). COSMO-CLM regional climate simulations in the Coordinated Regional Climate Downscaling Experiment (CORDEX) framework: a review. Geoscientific Model Development, 14(8), 5125–5154. https://doi.org/10.5194/gmd-14-5125-2021

Stevens, B., et al. (2013) The atmospheric component of the MPI-M Earth System Model: ECHAM6. J. Adv. Model. Earth Syst., https://doi.org/10.1002/jame.20015

Stouffer, R. J., Eyring, V., Meehl, G. A., Bony, S., Senior, C., Stevens, B., & Taylor, K. E. (2017). CMIP5 Scientific Gaps and Recommendations for CMIP6. Bulletin of the American Meteorological Society, 98(1), 95–105. https://doi.org/10.1175/BAMS-D-15-00013.1

Stratton, R. A., Senior, C. A., Vosper, S. B., Folwell, S. S., Boutle, I. A., Earnshaw, P. D., … Wilkinson, J. M. (2018). A Pan-African Convection-Permitting Regional Climate Simulation with the Met Office Unified Model: CP4-Africa. Journal of Climate, 31(9), 3485–3508. https://doi.org/10.1175/JCLI-D-17-0503.1

Sun Q., Miao C., Hanel M., Borthwick A.G., Duan Q., Ji D., Li H. (2019). Global heat stress on health, wildfires, and agricultural crops under different levels of climate warming. Environ. Int., 128,, 125-136

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). Geoscientific Model Development. https://doi.org/10.5194/gmd-12-4823-2019

Takaya, Y., et al. (2018). Japan Meteorological Agency/Meteorological Research Institute-Coupled Prediction System version 2 (JMA/MRI-CPS2): atmosphere–land–ocean–sea ice coupled prediction system for operational seasonal forecasting. Clim Dyn 50, 751–765. https://doi.org/10.1007/s00382-017-3638-5

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'ishi, R., Yamazaki, D., Suzuki, T., Kurogi, M., Kataoka, T., Watanabe, M., and Kimoto, M. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. Geoscientific Model Development. https://doi.org/10.5194/gmd-12-2727-2019

Taylor K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res-Atmos 106(D7):7183–7192

Taylor, K. E., Stouffer, R. J. and Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. Bull. Amer. Met. Soc., 93, 485-498.

Tebaldi C., Debeire K., Eyring V., Fischer E., Fyfe J., Friedlingstein P., Knutti R., Lowe J., O'Neill B., Sanderson B., van Vuuren D., Riahi K., Meinshausen M., Nicholls Z., Tokarska K.B., Hurtt G., Kriegler E., Lamarque J.-F., Meehl G. et al (2021). Climate model projections from the scenario model intercomparison project (Scenariomip) of cmip6. *Earth Syst Dyn* 12 253–293. https:// doi. org/ 10. 5194/ esd- 12- 253- 2021.

Tian, T., Yang, S., Karami, M. P., Massonnet, F., Kruschke, T., and Koenigk, T. (2021). Benefits of sea ice initialization for the interannual-to-decadal climate prediction skill in the Arctic in EC-Earth3. Geoscientific Model Development. https://doi.org/10.5194/gmd-14-4283-2021

Tong, Y., Gao X., Han Z., Xu Y. & Giorgi, F. (2020). Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods, *Climate Dynamics*, 57, 1425–1443.

Torma, C., Giorgi, F., & Coppola, E. (2015). Added value of regional climate modeling over areas characterized by complex terrain-Precipitation over the Alps. Journal of Geophysical Research: Atmospheres, 120(9), 3957–3972. https://doi.org/10.1002/2014JD022781

Torralba, V., Doblas-Reyes, F. J., MacLeod, D., Christel, I., and Davis, M. (2017). Seasonal climate prediction: A new source of information for the management of wind energy resources, J. Appl. Meteorol. Climatol., 56(5), 1231–1247, https://doi.org/10.1175/JAMC-D-16-0204.1

Vajda, A., and O. Hyvärinen (2020). Development of seasonal climate outlooks for agriculture in Finland. Adv. Sci. Res., 17, 269–277, https://doi.org/10.5194/asr-17-269-2020.

Van Schaeybroeck, B. and Vannitsem, S. (2011). Post-processing through linear regression, Nonlinear Process. Geophys., 18(2), 147–160, https://doi.org/10.5194/npg-18-147-2011

Van Schaeybroeck, B. and Vannitsem, S. (2015). Ensemble post-processing using member-by-member approaches: theoretical aspects, Q. J. R. Meteorol. Soc., 141(688), 807–818, https://doi.org/10.1002/qj.2397

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al. (2019) Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. J Adv Model Earth Syst 11(7):2177–2213. https://doi.org/10.1029/2019MS001683

Von Storch, H. and Zwiers, F. W. (2001). Statistical Analysis in Climate Research. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511612336

Vrac, M., Thao, S., & Yiou, P. (2022). Should Multivariate Bias Corrections of Climate Simulations Account for Changes of Rank Correlation Over Time?. *Journal of Geophysical Research: Atmospheres*, *127*(14), e2022JD036562.

Wainwright, C. M., Black, E., & Allan, R. P. (2021). Future Changes in Wet and Dry Season Characteristics in CMIP5 and CMIP6 Simulations, Journal of Hydrometeorology, 22(9), 2339-2357. Retrieved Jul 25, 2022, from https://journals.ametsoc.org/view/journals/hydr/22/9/JHM-D-21-0017.1.xml

Wainwright, Caroline & Allan, Richard & Black, Emily. (2017). Identification of deficiencies in seasonal rainfall simulated by CMIP5 climate models. Environmental Research Letters. 12. 10.1088/1748-9326/aa869e.

White, R. H., Toumi, R. (2013). The limitations of bias correcting regional climate model inputs. *Geophysical Research Letters*, 40, 2907-2912.

Wilks, D. S. (2011). Statistical Methods in the Atmospheric Sciences. International Geophysics Series, Academic Press. Volume 100, pp. 301-394. https://doi.org/10.1016/B978-0-12-385022-5.00008-7

Williams, K. D., et al. (2018). The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. Journal of Advances in Modeling Earth Systems, 10(2), 357– 380. https://doi.org/10.1002/2017MS001115

Willmott, C.J. (1981). On the validation of models. Phys. Geogr. 2, 184–194. https://doi.org/10.1080/02723646.1981.10642213

Wilson, A. Avila-Diaz, L.F. Oliveira, C.F. Zuluaga, B. Mark (2022). Climate extremes and their impacts on agriculture across the Eastern corn belt region of the U.S. Weather Clim. Extrem., 37, 100467, https://doi.org/10.1016/j.wace.2022.100467

WMO (2018). Guide to Competency. WMO No. 1205. World Meteorological Organization. Geneva: WMO. https://library.wmo.int/doc_num.php?explnum_id=4237

WMO (2019). State of the Climate in Africa. WMO-No. 1253. Geneva: World Meteorological Organization, ISBN 978-92-63-11253-8. https://library.wmo.int/doc_num.php?explnum_id=10421

WMO (2020). State of the Climate in Africa. WMO-No. 1275. Geneva: World Meteorological Organization, ISBN 978-92-63-11275-0. https://library.wmo.int/doc_num.php?explnum_id=10929

Wu, T., Zhang, F., Zhang, J., Jie, W., Zhang, Y., Wu, F., Li, L., Yan, J., Liu, X., Lu, X., Tan, H., Zhang, L., Wang, J., and Hu, A. (2020). Beijing Climate Center Earth System Model version 1, (BCC-ESM1): model description and evaluation of aerosol simulations, Geosci. Model Dev., 13, 977–1005, https://doi.org/10.5194/gmd-13-977-2020.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., et al. (2019). The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6. Geoscientific Model Development. https://doi.org/10.5194/gmd-12-1573-2019

Yeager, S. G., Danabasoglu, G., Rosenbloom, N., Strand, W., Bates, S., Meehl, G., et al. (2018). Predicting near-term changes in the Earth System: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. Bulletin of the American Meteorological Society. https://doi.org/10.1175/BAMS-D-17-0098.1

Yu, K., Hui, P., Zhou, W., & Tang, J. (2021). Evaluation of extreme temperature in multi-RCM simulations over CORDEX-East Asia phase II domain. Atmospheric Research, 255, 105535. https://doi.org/10.1016/j.atmosres.2021.105535

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa., S., et al. (2019). 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. https://doi.org/10.2151/jmsj.2019-051

Zhu, X., & Troy, T. J. (2018). Agriculturallyrelevant climate extremes and theirtrends in the world's major growingregions.Earth's Future,6, 656–672.https://doi.org/10.1002/2017EF000687