

**Developing gridded climate data using neural networks: high-resolution historical
climate and future projections for Africa**

by

Sarah Aminah Namiiro

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Forest Biology and Management

Department of Renewable Resources

University of Alberta

© Sarah Aminah Namiiro, 2024

Abstract

Databases of high-resolution interpolated climate data are essential for climate change research, such as analyzing impacts of climate extreme events on biological systems and the development of climate change adaptation strategies for managed and natural ecosystems. To enable such efforts, this thesis contributes a comprehensive high-resolution database of historical and future climate data for Africa, including 30-year normal period estimates, decadal averages, annual, seasonal, and monthly data from 1901-2020, as well as CMIP6 multi-model climate projections for the 2020s, 2050s, and 2080s.

The database contains 48 monthly variables (Tmin, Tmax, Tave, Prec), 16 seasonal variables, and 16 bioclimatic variables (such as degree days, drought indices, etc.) for 142 historical time periods (years, decades, normals) and 168 future projections. A collaborator provided a user-friendly software solution (ClimateAF), with database and the software front-end freely available for download at <http://tinyurl.com/ClimateAF>, allowing non-technical users to interactively query a total of 24,800 climate grids generated in this study.

These climate grids are queried with the delta (change factor) method, which is based on a high-quality baseline grid for the 1961-1990 normal period (2.5 arcminute resolution), while all other grids are derived by expressing historical and future data as a lower resolution anomaly layer (0.5 degree), minimizing data storage requirements. The ClimateAF software combines the layers and then further downscales to any desired resolution (up to 250m practically useful resolution in mountainous terrain) with empirical environmental lapse rates.

The 1961-1990 baseline grid for Africa was developed using a combination of thin-plate spline interpolation of data from 4625 weather stations for Africa, and subsequent fine-tuning

with neural networks that associate climate observed at weather stations with covariates reflecting topographic and geographic information (such as elevation, aspect, slope, distance to coast and lakes) in combination atmospheric data from the ERA5 general circulation model (monthly wind direction and strength). The resulting baseline grids accurately model climate phenomena such as precipitation induced by orographic lift on the windward side of mountains, rain shadows, and lake effects on temperature in their vicinity.

Future climate projections were obtained from 13 Atmosphere-Ocean General Circulation Models (AOGCMs) of the sixth phase of the Coupled Model Intercomparison Project (CMIP6) for four emission scenarios (SSP 1-2.6, 2-4.5, 3-7.0 and 5-8.5) and three future time periods (2020s, 2050s and 2080s). To support users in the selection of a representative set of scenarios for different regions of Africa, I used the Katsavounidis-Kuo-Zhang (KKZ) algorithm which selects an optimally representative set of future projections for 11 regions of Africa, given a user-requested number of scenarios.

The database was validated using two approaches: To optimize thin-plate spline models and neural network fine tuning, I used a checkerboard validation approach, where the study area was divided into three degree grid cells. Half the cells (“black fields”) were used for model training and the other half (“white fields”) for validation, and vice versa. This approach controls spatial autocorrelations among nearby weather stations and thereby avoids overfitting. The final model was built using all data, and error statistics were assessed using the mean absolute error (MAE) in a non-independent test, including the additional accuracy improvements from ClimateAF’s downscaling algorithms.

Preface

This thesis is being prepared for submission as a journal article. Additional contributors to this publication are Andreas Hamann, Tongli Wang, Dante Castellanos-Acuna and Colin Mahony. The study was conceived by AH and TW. I compiled and curated the weather station data, and generated the gridded climate data with advice from AH. DCA contributed covariates for model development, TW developed the ClimateAF software front-end to query the database, and CM provided the algorithm for multi-model ensemble selection. I generated all figures and tables and wrote a first draft of the thesis, reviewed and edited by AH.

Acknowledgements

My entire period of study at the University of Alberta has been worthwhile and very enlightening. Particularly, the process of this study and writing this thesis has been a tremendous achievement and exhilarating learning experience. I could not have accomplished this without the great support system I had.

First, with my deepest gratitude, I thank Dr. Andreas Hamann, who has been my supervisor and mentor. His belief and encouragement has led me to some of the most rewarding moments in my life including undertaking this study, teaching assistantship and several awards. His brilliance and enthusiasm have been a constant source of inspiration throughout this study. As an international student, he has also made me feel most at home and I can't thank him enough for that.

Second, I thank my parents and family for their unwavering support and encouragement. Even distances apart, their love and guidance has upheld me. I especially thank my dad Zaake H. Mukasa, for making me feel like I can achieve anything. To my nieces and nephews, you bring joy and warmth to my life every day.

To my friends especially the members of the SIS Lab, you have been tremendous company and support. I admire your brilliance and I believe you are all destined for great things. I thank you for being friendly and availing a cheery and conducive space in which I've thrived. A special mention to Dante CA who has particularly been an amazing and kind friend.

I would also like to thank Dr. Michael Mbogga at Makerere University who has been a key part of this journey from the start. His belief and support have led me here.

Finally, I thank the staff and faculty of UofA especially in the Department of Renewable Resources. I've learned a great deal from the interactions and engagements I've shared with you all.

I hope that in some way, you can all receive the kindness and gratitude you have extended to me.

Table of Contents

Abstract.....	ii
Preface.....	iv
Acknowledgements.....	v
Table of Contents.....	vi
List of Tables.....	viii
List of Figures.....	ix
List of Acronyms.....	xi
1. Introduction.....	1
2. Literature review.....	5
2.1 Climate variables and summaries.....	5
2.2 Approaches to generate interpolated climate data and available products.....	7
2.3 Artificial neural networks for climate spatial interpolation.....	11
2.4 Future climate, GCMs and ensemble selection.....	13
3. Data and Methods.....	16
3.1 Study Area.....	16
3.2 Weather station data compilation.....	18
3.3 Thin-plate spline interpolation.....	22
3.4 Covariates.....	22
3.5 Fine tuning interpolations using artificial neural networks.....	26
3.6 Checkerboard validation.....	28
3.7 CMIP6 models cluster analysis and regional ordered subset selection.....	30
4. Results.....	33
4.1 Interpolated baseline grids.....	33
4.2 Validation of climate estimates.....	38
4.3 Selection of representative CMIP6 models.....	40
5. Discussion.....	47
6. Conclusion.....	52
References.....	54
Appendix.....	80
A. Initial thin-plate spline interpolation with binned weather station data.....	80

B. Fine-tuning the thin-plate spline interpolations with a neural network 81

List of Tables

Table 1. Databases included in this study detailing their temporal extent, temporal resolution and the number of stations obtained for Africa. The number of stations in parentheses are the stations that were selected after quality control and duplicate removal.....	18
Table 2. Predictor variables (features) for training the neural network. The original target resolution was 2.5 arcminutes, and low-pass filters were applied to better predict larger scale climate patters driven by higher altitude air circulation patterns.....	24
Table 3. The final neural network model structure and parameter settings	28
Table 4. Models pre-selected by Mahony <i>et al.</i> (2022) in the 13-model ensemble with their institution and citation.....	31
Table 5. Cross-validation statistics for the ANN model. MAE is absolute mean error, R-squared is the coefficient of determination and RMSE is the root mean squared error. Tmin indicates monthly minimum temperature, Tmax is monthly maximum temperature and Prec is monthly precipitation.	39
Table 6. Mean absolute error (MAE) between observed and interpolated surfaces for monthly temperature variables (Tmin, Tmax, Tave) and precipitation (Prec), seasonal average temperature (Tave _s) and seasonal precipitation (Prec _s) and, annual temperature (MAT) and annual precipitation (MAP) aggregated for the 11 IPCC climate regions for Africa: ARP: Arabian-Peninsula; CAF: Central-Africa; ESAF: East-Southern Africa; MDG: Madagascar MED: Mediterranean; NEAF: North-Eastern-Africa; SAH: Sahara; SEAF: South-Eastern-Africa; WAF: Western-Africa (WAF); WCA: West central Asia and; WSAF: West-Southern-Africa. AF represents the continental average for Africa.	40
Table 7. 13-model ensemble projected means of mean annual temperature (MAT; °C) and mean annual precipitation (MAP; %) by the 11 IPCC regions (Figure 1) and AF (continental average) for the future scenarios; SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5. See Table 6 for region abbreviations.	41
Table 8: Subsets of the projections with optimal representation variation in climate change projections for a given subset size according to, including additional selection criteria from Mahony <i>et al.</i> (2022). For example, a 4-model ensemble for the ARP region would include CNRM-ESM2-1, UKESM1.0-LL, EC-Earth3 and, MPI-ESM1.2-HR. Users may choose to exclude the extreme projections of UKES as well. See the Table 6 caption for region abbreviations.	44
Table 9: Variable loadings of the PCA plots in Figure 11. Seasonal variables are for winter (wt; DJF), spring (sp; MAM), summer (sm; JJA) and autumn (at; SON) for the minimum monthly temperature (Tmin), maximum monthly temperature (Tmax), average monthly temperature (Tave) and precipitation (prec). Monthly maximum and minimum temperature for January to December were also used for the ordination.....	46

List of Figures

Figure 1: Africa’s IPCC climate reference regions and Köppen-Geiger climate classification (Main Climates: A – equatorial, B – arid, C –warm temperate, D – snow, E – polar; Precipitation: W – desert, S – steppe, f – fully humid, s – summer dry, w – winter dry, m – monsoonal; Temperature: h – hot arid, k – cold arid, a – hot summer, b – warm summer, c – cool summer, d – extremely continental, F – polar frost, T – polar tundra (open access data for this map was obtained from Kottek *et al.*, 2006; Iturbide *et al.*, 2020). 17

Figure 2: Distribution of 4625 weather stations compiled for the database. Blue stations have records for only precipitation measurements and red stations have records for both precipitation and temperature measurements for the 1961-1990 period..... 21

Figure 3: Example of predictor variables used in neural network fine-tuning of thin-plate spline interpolations. All putative predictor variables were subjected to transformations for normality if possible, and then scaled to values between 0 and 1 for use as covariates in neural network models. 25

Figure 4. 3° x 3° checkerboard grid over Africa for cross-validation groups. Black points are the weather station location falling in either the “white tile” for training or “black/gray tile” for testing..... 30

Figure 5: Example of an interpolated climate grid of January precipitation for 1961-1990. The inset highlights the influence of the micro-climatic patterns in the southeast African highlands and coast. The color of circles in the inset indicates the weather station values on the same scale. 34

Figure 6. Adjustment to the TPS interpolated climate grid for Prec01 for 1961-1990. The inset highlights adjustment to southeast African highlands. The color of circles in the inset indicates the weather station residuals (predicted minus observed) on the same scale. 35

Figure 7. Example of an interpolated climate grid for January minimum monthly temperature for 1961-1990. The inset highlights the influence of the micro-climatic patterns in the southeast African highlands. The color of circles in the inset indicates the weather station values on the same scale. 36

Figure 8. Interpolation adjustment by the ANN to the TPS interpolation for Tmin01. The inset is a close-up of the adjustment in the southeastern Africa around Lake Malawi. The color of circles in the inset indicates the weather station residuals (predicted minus observed) on the same scale. 37

Figure 9. Feature importance for neural network interpolation for the January monthly variables average minimum temperature (Tmin01), average maximum temperature (Tmax01) and precipitation (Prec01). The candidate covariates are provided at various resolutions with low-pass filters (e.g., an “lp5” extension indicates a 5x5 grid cell average for 2.5 arcminute grid cells). 38

Figure 10: Models are structured by a cluster dendrogram showing spatial similarity in the projected seasonal changes for Tmin, Tmax and precipitation in winter (DJF) and spring (MAM) in the period 2041-2070 under (SSP2-4.5). The maps illustrate the visual changes across the African continent for this period. Precipitation is log-scaled to provide proportional magnitude of positive and negative changes..... 43

Figure 11: Illustration of the selection process by the KKZ algorithm for the region of west-southern Africa (WSAF) as an example. The first model is selected based on the ensemble centroid in multivariate space, represented here as principal component scores of climate variables (a). The four quadrants (brown labels in panel a) represent primarily temperature gradients (PC1) and precipitation gradients (PC2), with the corresponding variable loadings of the PCA provided in Table 9. The panels (b) to (d) show the sequential selection of additional scenarios to optimally represent uncertainty with given a minimum number of scenarios, where the multivariate Mahalanobis distances are shown with red numbers..... 45

Figure 12: Comparison of interpolated climate grids for January precipitation, fine-tuned with neural networks from this study (a), with a Google Earth satellite image (b) and other popular and well-regarded data products: WorldClim (c), CHELSA (d). Weather station data is represented by open circles with fill values on the same scale as the climate interpolations..... 47

List of Acronyms

AGCD	Australian Gridded Climate Data
AHM	annual heat-to-moisture index
ANN	Artificial Neural Network
ANUSPLIN	Australian National University Spline
AOGCM	Atmosphere-Ocean General Circulation Models
AR	Assessment report of the IPCC
ARC2	Africa Rainfall Climatology for famine early warning systems
CART	Classification and regression trees
CHELSA	Climatologies at high resolution for the earth's land surface areas
CHIRPS	Climate Hazards Group InfraRed Precipitation
CHIRTS	Climate Hazards center InfraRed Temperature
CMIP6	Coupled Model Intercomparison Project phase 6
CORDEX	Coordinated Regional Climate Downscaling Experiment
CRU	Climatic Research Unit
CTI	Compound topographic index
DEM	Digital elevation model
DL	Deep learning
DNN	Deep neural networks
ECA&D	European Climate Assessment and Dataset
ECS	Equilibrium climate sensitivity
ENSO	El Nino-southern oscillation
ERA5	European Centre for Medium-Range Weather Forecasts Reanalysis v5
ESRI	Environmental Systems Research Institute
FAO	Food and Agricultural Organization of the United Nations
GHCN	Global Historical Climatology Network
IDW	Inverse distance weighting
IPCC	Intergovernmental Panel on Climate Change
ITCZ	Inter-tropical convergence zone
KKZ	Katsavounidis-Kuo-Zhang algorithm
MAE	Mean absolute error
MAP	Mean annual precipitation
MAT	Mean annual temperature
MCMT	Mean coldest-month temperature
ML	Machine learning
MSE	Mean square error
MWMT	Mean warmest-month temperature
NN	Neural network
NOAA	National oceanic & atmospheric administration
OK	Ordinary kriging

PCMDI	Program for Climate Model Diagnosis and Intercomparison
Prec	Monthly precipitation
PRISM	Parameter-elevation Relationships on Independent Slopes Model
R	R programming language
ReLU	Rectified linear unit
RF	Randomforest
RMSE	Root mean squared error
SGD	Stochastic gradient descent
SSP	Shared socioeconomic pathways
Tave	Monthly average temperature
Tmax	Monthly maximum temperature
Tmin	Monthly minimum temperature
TPI	Topographic position index
TPS	Thin plate splines
UK	Universal kriging
WMO	World meteorological organization

1. Introduction

With the increasing concern over climate change, interpolated climate data have become essential for assessing climate change impacts, monitoring natural and managed ecosystems, and designing conservation and adaptation strategies to minimize climate change impacts (Hewitt *et al.*, 2020; Omukuti *et al.*, 2023). Virtually every study in the field of climate change impacts and adaptation requires information on long-term climate conditions (30-year climate normals), records of past climate variability (monthly, seasonal, and annual historical data), as well as future projections from Atmosphere-Ocean General Circulation Models (AOGCM). Such data are now widely available at global scales, but they vary widely in quality among continents and regions depending on weather station coverage, as well as in the methodological approaches that were used to generate interpolated climate grids.

Gridded climate products can be broadly classified into three categories based on methodologies used to generate them. The first approach is direct interpolation of ground-based weather station data using techniques such as kriging, inverse distance weighting and splines. Examples include widely used products such as WorldClim (Hijmans *et al.*, 2005; Fick & Hijmans, 2017), Climatic Research Unit gridded Time Series (CRU-TS) grids (New *et al.*, 1999; Harris *et al.*, 2014). The second approach uses Atmosphere-Ocean General Circulation Models (AOGCM) run on historical weather station data, the same models that provide daily and hourly weather forecasts. Gridded historical climate data generated with this approach is referred to as a reanalysis product. Examples for widely used global reanalysis products are CHELSA (Climatologies at high resolution for the earth's land surface areas) (Karger *et al.*, 2017) and ERA5 (European Centre for Medium-Range Weather Forecasts Reanalysis v5) (Hersbach *et al.*, 2020). These datasets offer global coverage with spatial resolutions to as high as a kilometer and at daily temporal resolution (but not in combination). The third approach uses satellite-based remote sensing, calibrated with ground-based weather station data, to estimate precipitation and temperature. Examples include the global Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) (Funk *et al.*, 2014; Funk *et al.*, 2015) and the temperature counterpart, CHIRTS (Verdin *et al.*, 2020).

For some regions of the world, there are also well-regarded expert products that utilize multiple approaches to optimize climate estimates. For example the European Climate Assessment and Dataset (ECA&D) for Europe and surrounding areas fits station data in a two stage process including a thin-plate splines (TPS) and bilinear interpolation of residuals treated under Gaussian Random Field model to produce the daily estimates (Cornes *et al.*, 2018). The Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE) dataset provides daily gridded data for Asia including the Himalayas, and the mountainous regions of the Middle East (Yatagai *et al.*, 2010; Yatagai *et al.*, 2012). A particularly well-regarded regional product is the Parameter-elevation Relationships on Independent Slopes Model (PRISM) that utilizes topographic information to model small-scale weather patterns such as orographic lift and rain shadows for the United States (Daly *et al.*, 2008). For Africa, regional comprehensive climate databases are lacking, although some high quality products exist for specialized applications, including the Africa Rainfall Climatology for famine early warning systems (ARC2) (Novella & Thiaw, 2013), the University of Reading's TAMSAT African Rainfall Climatology And Time series (TARCAT) (Maidment *et al.*, 2014). Both ARC2 and TARCAT combine satellite imagery and interpolation of climate station data to obtain high quality precipitation grids.

Both global and regional products make compromises in the extent of their spatial and temporal resolution, climate variable coverage, inclusion of future projections, and coverage of historical time periods for computational reasons and data management limitations. To address these limitations, Wang *et al.* (2016), Wang *et al.* (2012) and Marchi *et al.* (2020) developed software packages for North America and Europe that build on a high-quality, high-resolution 1961-1990 climate baseline, developed by Daly *et al.* (2008) with the Parameter-elevation Relationships on Independent Slopes Model (PRISM). Climate grids for historical and future periods of comparable quality are then generated by the software using the delta (change factor) method by overlaying lower resolution (0.5 degree) anomaly layers, minimizing data storage requirements. As the last step, the software downscales from an internal 2.5 arcminute (approx. 4 km) internal grid resolution to any user provided GPS coordinate or user-provided digital elevation model. The downscaling approach is driven by empirical environmental lapse rates that correlate local climate gradients with elevation gradients.

To develop an equivalent software package for Africa that allows easy access to a variety of climate grids, three components are needed: (1) historical climate data to develop past climate normal periods, decadal averages, annual, seasonal and monthly data from 1901 to present, (2) CMIP6 multi-model climate projections for 20-year and 30-year future periods, and (3) a high-quality, high-resolution baseline grid for a reference normal period. Low resolution historical anomaly data is readily available globally from several sources, and I have selected the well regarded CRU-TS time series data for its available data coverage from 1901 to present. An advantage of this dataset is that it interpolates anomalies, and approximates the 1961-1990 normal period for regions that lack weather station data, which are prevalent in some regions of Africa, especially in the early 19th century. Future climate projections generated by AOGCMs are also readily available as a consolidated product under the sixth phase of the Coupled Model Intercomparison Project (CMIP6) (PCMDI, 2022).

However, a high-resolution, high-quality reference normal period is not available for Africa. PRISM baseline data is not available outside of North America and Europe, and the PRISM methodology is an expert system that is not openly accessible. A key objective of my thesis research is therefore to develop an alternative interpolation approach that is capable of modeling local weather patterns in complex topographic terrain, such as rain shadows, orographic lift, temperature inversions, topographic shadowing, or lake effects. A promising methodological approach to model complex local climate patterns that are non-linear and the result of multiple interacting topographic and atmospheric factors are artificial neural networks (Antonić *et al.*, 2001; Rampal *et al.*, 2022), which I will test for their utility to develop a 1961-1990 climate normal reference grid for Africa.

Another objective of this thesis project is to address a common issue faced by users in the selection of a representative set of future climate scenarios. The CMIP6 dataset includes over 120 AOGCMs (Eyring *et al.*, 2016; PCMDI, 2024) that are at varying quality for the different temporal extents including numerous scenarios of shared socio-economic pathways (SSPs) of future climate (Meinshausen *et al.*, 2020). This results in a wide range of projections for every locality, making general recommendations for use of specific AOGCM projections difficult (Knutti *et al.*, 2010). Conversely, averaging multiple models or selecting a single median model to represent probable future climatic conditions is also not recommended for lack of capturing

the uncertainty in current future projections (Endris *et al.*, 2013; Whittleston *et al.*, 2017). To address this problem, I build on work by Mahony *et al.* (2022), who pre-selected 13 CMIP6 models for desirable attributes, including model complexity, resolution, variable coverage, bias correction and computational investments in multiple model runs. Starting with the 13 pre-selected AOGCMs, I will then use the Katsavounidis-Kuo-Zhang (KKZ) algorithm, also following Mahony *et al.* (2022) to select future projections that maximize representation of uncertainty for a given number of scenarios for different regions in Africa. For example, the user can request three scenarios for sub-Saharan western Africa, and the algorithm will select three scenarios that best represent likely future climates, avoiding outliers and duplicate selection of similar scenarios.

In summary, this thesis aims at developing a comprehensive, high-resolution database of historical climate and future projections for Africa. My specific objectives are to (1) develop a comprehensive monthly weather station database for Africa for interpolation and validation, (2) develop and validate interpolated climate grids for Africa using machine learning with neural networks, and (3) provide guidance for selecting multi-model ensemble subsets of CMIP6 models for Africa's IPCC regions.

2. Literature review

2.1 Climate variables and summaries

Climate data is primarily obtained from weather station records. Common weather variables recorded include daily minimum temperature, daily maximum temperature, precipitation, humidity, wind speed and direction, and atmospheric pressure, among other less frequently reported statistics, such as sunshine hours, snow depth, etc. Weather and climate conditions are then summarized as hourly, daily, monthly, seasonal, annual, decadal and 30-year averages, the latter being commonly referred to as climate normals for a specific period, such as a month, season or year. Climate normals, such as January temperature or summer precipitation, serve as an expectation of average weather conditions (Smith, 1975; Guttman, 1989). Conversely, the finest temporal resolution can be important to track the occurrence of extreme weather events (Hartmann *et al.*, 2013). Additionally, some weather variables are best expressed with metrics representing their diurnal cycles for example wind speed or maximum and minimum daily temperature (Wilks, 2019).

For short-term planning and decision-making, such as agricultural management or disaster preparedness, monthly or seasonal summaries may be more appropriate. Monthly variables also offer a concession between managing large amounts of daily data variables and apprehending seasonal climate variation (Castellanos-Acuna & Hamann, 2020). Annual variables are important to track inter-annual variability and particularly useful in some models such as the hidden state Markov (HSM) model that can ascertain drought risk (Thyer & Kuczera, 2000). Additionally annual values can also be disaggregated to monthly (Mejia & Rousselle, 1976; Tarboton *et al.*, 1998) or even daily values (Wilks, 2019). On the other hand, long-term climate assessments and policy formulation require summaries at decadal or longer time periods scale to provide a more comprehensive understanding of climate patterns and variability (Arguez & Vose, 2011).

Documenting weather summaries over a span of a thirty-year period (climate normal period), enables the computation of an average that infers the long-term climate expectation of a given region for that normal period (Guttman, 1989). These averages are usually not biased by random and cyclical anomalies and are useful for comparisons (Guttman, 1989) and predictions of climate change in environmental and other sectoral studies (Arguez & Vose, 2011). The 1961-

1990 climate normal period has significant popularity because of its prominence in representing pre-industrial climate in various studies (Engelbrecht *et al.*, 2015; Marchi *et al.*, 2019).

Future climate projections are usually done on near-term or long-term basis. Near-term projections focus on the period from the present to mid-century (2050) and use various summaries including annual, multi-year and decadal (typically 10 and 20 years). For example a 10-year decadal prediction (2011-20) over North America was done by (Hoerling *et al.*, 2011) and the assessment reports of IPCC, the fifth (AR5) and the sixth (AR6) use 20 year average 2016-2035 (Kirtman *et al.*, 2013) and 2021-2040 (Lee *et al.*, 2021) respectively. Near-term climate predictions are particularly important to decision makers in government and industry (Kirtman *et al.*, 2013). Further into the future, summaries are typically over decades for instance the mid-term period in the AR6 uses the 20-year average (2041-2060) and the long-term period in AR5 and AR6 is 30-year average (2081-2100) (Collins *et al.*, 2013; Lee *et al.*, 2021). Near-term projections diverge from longer term projections as they are less sensitive to differences between future emission scenarios and are instead more sensitive to short-lived climate forcing agents such as volcanic activities, substantial changes in solar irradiance, land use changes (Lee *et al.*, 2006; Räisänen & Ruokolainen, 2006; Kirtman *et al.*, 2013).

In addition to the primary variables; monthly minimum temperature, monthly maximum temperature and precipitation, other variables sometimes referred to as bioclimatic variables can be derived from them to provide more ecologically meaningful representation of climate (Booth, 1985; Deblauwe *et al.*, 2016). These derived variables can either be directly calculated or computed by equations from daily or monthly data (Wang *et al.*, 2012). Directly derived variables include mean annual temperature (MAT), mean coldest-month temperature (MCMT), mean warmest-month temperature (MWMT), continentality (difference between MWMT and MCMT), mean annual precipitation (MAP), annual heat-to-moisture index (AHM) and seasonal variables (Wang *et al.*, 2006; Wang *et al.*, 2012). Equation derived variables such as degree-days and frost-free days, precipitation as snowfall among others usually estimated from daily data have successfully been obtained from monthly data or using both (Wang *et al.*, 2006; Schlenker *et al.*, 2007). These derived variables have been used in genetic, species distribution and ecological modelling and studies in different parts of the world (Cortini *et al.*, 2012; Bradley St. Clair *et al.*, 2013; De La Torre *et al.*, 2014; Mathur *et al.*, 2023). In Atlantic Central Africa,

Deblauwe *et al.* (2016) studied the species-climate association of angiosperm plants using temperature and precipitation derived variables. Species distribution modelling especially in the advent of climate change has become common in Africa; for example, Ngarega *et al.* (2023) used temperature seasonality, maximum temperature of the warmest month and other variables to determine the potential distribution of *Lippia javanica* in tropical and sub-tropical Africa and, Ogougbé *et al.* (2022) used a number of bioclimatic variables to model the ecological niche of *Harrisonia abyssinica* in West Africa in future climates.

2.2 Approaches to generate interpolated climate data and available products

Weather stations records are not available for all geographical locations for various reasons including limited finance into weather and climate information and infrastructure (Georgeson *et al.*, 2017), bias towards areas of high population, and often neglect areas such as mountains and deserts that pose additional logistical and accessibility challenges (New *et al.*, 1999; Menne *et al.*, 2012). To obtain complete spatial coverage, weather station data is interpolated (smoothed) using various techniques that use mathematical algorithms to estimate missing data from observed values and other variables. The accuracy of interpolated surfaces is reliant on the accuracy of input data, spatial variability and the algorithm used (Hartkamp *et al.*, 1999). Interpolation techniques range from the simplistic methods such as use of nearest neighbors to more complex, deterministic methods such as inverse distance weighting (IDW) and Thiessen polygons, stochastic methods such as kriging, splines and machine learning methods such as random forest and neural networks (Hartkamp *et al.*, 1999; Rigol *et al.*, 2001).

Kriging is one of the common methods used in climatic interpolation. The kriging algorithms basically predict the value of a variable at an unobserved location by using weighted averages of observed values, based on the spatial autocorrelation modeled by a semivariogram. The algorithms rely on Gaussian processes that minimize variance (Matheron, 1963; Cressie, 1993). A popular variant of kriging is the ordinary kriging (OK) (Matheron, 1971) and was used in combination with nearest neighbors to model monthly mean air temperature for Brazil (Alvares *et al.*, 2013). However, Cressie (1993) cautioned that OK is highly influenced by large changes or outliers making it prone to anomalous overestimations. A variant to address this problem was robust kriging (Hawkins & Cressie, 1984) that uses neighboring values to determine how an

outlier may be down-weighted. Other variants include universal kriging (UK), median-polish kriging, simple kriging and kriging with external drift (Cressie, 1993; Hoaglin *et al.*, 2000; Chiles & Delfiner, 2012). Wu and Li (2013) developed and used residual kriging to interpolate monthly average temperature in the United States. Sub-regional grids for West Africa were developed by (Bliefernicht *et al.*, 2022) using station data from GHCN and Global Surface Summary of Day (GSOD) and interpolation by OK. Kriging was reported to be advantageous when interpolating with few known values because it assigns a variance to all known and missing values making the regression robust (Shrestha *et al.*, 1999). However, this was found to be a potential limiting factor for kriging methods to perform low-bias interpolations (Nalder & Wein, 1998).

IDW interpolates unknown values based on the inverse distance of their location to the point of nearby known values (Shepard, 1968). Attorre *et al.* (2007) reviewed the performance of detrended-IDW in interpolating climatic and bioclimatic variables where it outperformed UK and a neural network for one of 21 variables interpolated. In another study, the gradient-plus-inverse distance squared (GIDS) technique, which combines multiple linear regression and distance-weighting outperformed several kriging methods in the interpolation in climate normals for monthly temperature and precipitation in western Canada (Nalder & Wein, 1998). Distance interpolation techniques are opted for over other methods because of the simplicity of their functions and in the earlier years for their low computational demands on computers (Attorre *et al.*, 2007).

Splines interpolate by fitting a smooth curve to a set of data points, allowing for the estimation of values at un-sampled locations (Wahba, 1979). Variants include linear and cubic splines based on the polynomials utilized for interpolation (Elhakeem *et al.*, 2022). A highly used variant in climate studies are the thin plate splines (TPS) that use non-polynomial functions allowing for more flexibility and higher levels of smoothness (Hutchinson & Gessler, 1994). TPS have been used to interpolate mean rainfall globally by (Hutchinson, 1995a), obtaining low estimation error percentages (3-7%). In addition to being computationally efficient, TPS have an efficient detection of erroneous data (location and elevation) as it indicates large outliers from the fitted spline surface (Hutchinson, 1995a). Several studies have used TPS for interpolations including Jones *et al.* (1999) for surface air temperatures for global land areas using location and elevation

variables, Mbogga *et al.* (2010) for historic climate data for Alberta, Wang *et al.* (2012) for historic climate data for parts of western North America, van Niekerk and Joubert (2011) for western Cape Province in South Africa, among others. Some of these studies rely on the use of the convenient software ANUSPLIN (Australian National University Spline) that employs TPS for interpolation (Hutchinson & Xu, 2013). According to Hutchinson and Gessler (1994), one of the differences between splines and kriging is that splines are defined by minimizing the roughness of the interpolated surface, subject to having a prescribed residual from the data while kriged surfaces are defined by minimizing the variance of the error of estimation.

Machine learning (ML) methods including probabilistic modelling, kernel methods and classification and regression trees (CARTS) also known as decision trees including RandomForests (RF) and gradient boosting machines (GBM) and neural networks (NN) are increasingly being used in interpolation. One of the prominent techniques is RF which interpolates by randomly selecting and aggregating several CARTS for which a random subset of features are selected at each decision node (Breiman, 2001). The final prediction is represented by the average of the prediction from the individual trees in the “forest” (Breiman, 2001). RF has several variants including random forest for spatial predictions (RFsp) (Hengl *et al.*, 2018) and random forest spatial interpolation (RFSI) (Sekulić *et al.*, 2020). The technique has been used in several studies including the downscaling of mean temperature in southern China (Pang *et al.*, 2017) and estimation of precipitation in Switzerland (Wolfensberger *et al.*, 2021). In another study by Tan *et al.* (2021), they combined RF and IDW to generate climate surfaces of precipitation and temperature for China. The RF-IDW had similar accuracy as using RF alone for temperature but improved for the interpolation of precipitation. The advantages to RF is its excellence in managing large datasets and the inbuilt function to evaluate variable importance (Pang *et al.*, 2017). Neural networks, a ML method have also been used and reviewed by several studies (Antonić *et al.*, 2001; Attorre *et al.*, 2007; Zhan *et al.*, 2023) and are discussed in the next section of literature review.

Other notable interpolation techniques include PRISM by Daly *et al.* (2008) from Oregon University. The method relies on regression-based modelling to interpolate climate variables in physio-graphically complex landscapes. PRISM has been used to develop high quality climate surfaces for the US and other parts of North America. It was noted by Hutchinson (1995a) that

PRISM does not yield a continuous surface as do the TPS, a gap filled by post hoc filtering and other adjustments in the PRISM. Bilinear interpolation which offers simplicity using linear regression was used by (Wang *et al.*, 2016) for interpolation of climate surfaces in North America. Other studies incorporate other techniques to improve interpolation; for example, the Real-time hourly rapid update cycle (RUC) data used by (DeGaetano & Belcher, 2007). In a study by (Workneh *et al.*, 2024), the use of orographic-based linear interpolation with elevation, slope and aspect variables outperformed IDW and kriging methods in the interpolation of precipitation in Ethiopia.

Widely used gridded data such as WorldClim (Hijmans *et al.*, 2005; Fick & Hijmans, 2017) and CRU-TS grids to represent historical time series (New *et al.*, 1999; Harris *et al.*, 2014) were developed with thin-plate spline methodologies with just three predictor variables (latitude, longitude, and elevation), because they are computationally efficient and well established. More advanced products couple AOGCM re-analysis and remote sensing techniques with interpolation to improve their products. For example CHELSA (Climatologies at high resolution for the earth's land surface areas) use GCMs and B-splines (Karger *et al.*, 2017) and ERA5 (ECMWF Reanalysis v5) employs four-dimensional data assimilation for optimal interpolation (Hersbach *et al.*, 2020). Regional grids for Africa i.e. CHIRPS (Funk *et al.*, 2015), CHIRTS (Verdin *et al.*, 2020), ARC2 (Novella & Thiaw, 2013) were developed with a combination of satellite imagery and interpolation, mainly with IDW. One common drawback of gridded climate databases is their substantial size, making the process of extracting pertinent data for specific sample points or local areas of interest both laborious and necessitating proficient skills in geographic information systems (GIS). Additionally, these databases encounter inherent limitations when characterizing sample points with imprecise spatial information, such as those reported to the nearest minute. This issue becomes especially pronounced in regions characterized by steep mountainous terrain, where even a medium-resolution grid cell may encompass climate variations spanning several hundred meters in elevation (Wang *et al.*, 2016).

The choice of method used for interpolation can be influenced by a number of reasons and trade-offs among the model features. For example a technique may be chosen for its model performance (extent of fit), model complexity (ease of interpretation and acquiring estimates), spatial dependence of data, level of influence of local features (Attorre *et al.*, 2007), performance

under different spatial extents and data density (number of available weather stations) (Antonić *et al.*, 2001), computational efficiency among others. The accuracy of interpolation estimates also rely on the type of climate variable estimated. Selection of techniques may also be limited by the availability of software and computational demands (Li & Heap, 2014). Interpolated surfaces have been improved over the years by incorporating other variables besides station location data values, such as elevation (Willmott & Matsuura, 1995; Boer *et al.*, 2001), slope gradient, slope aspect, hill shade (van Niekerk & Joubert, 2011) to improve accuracy of interpolated surfaces.

2.3 Artificial neural networks for climate spatial interpolation

Deep learning and neural networks, a subfield in ML is gaining prominence in spatial interpolation (Li & Heap, 2014) for its practically unlimited ability to model non-linear relationships and high-level interactions in huge, multi-dimensional predictor variable datasets. Deep learning (DL) specially utilizes successive layers to develop meaningful representations learned via artificial neural networks (ANNs), their conceptual model inspired by nervous activity in the human brain (McCulloch & Pitts, 1990). They consist of interconnected nodes, or artificial neurons, organized in layers: input layer, hidden layers (if any), and output layer. Each neuron receives input signals, processes them, and produces an output signal. The ANN is trained by exposure to known examples of inputs and outputs in this case, observed weather station data. The data may be reiterated over the model a number of times i.e. “training” and the ANN “learns” the useful patterns, representations and rules in the input data within a predefined space and guidance from a feedback signal (Chollet, 2022). Neural networks (NNs) can range from a straightforward linear stack of layers to a sophisticated multi-head network. With more hidden layers and complexity, an ANN becomes a deep neural network (DNN). DNNs typically have higher computational demands and equally higher excellence in learning intricate features and representations in raw data. Neural network also differ by layer type. For instance, vector data is usually treated under dense layers, while sequence data is usually processed under recurrent or convolutional layers, hence the names recurrent neural networks (RNN) or convolutional neural networks (CNN). Image data is further processed by more complex 2D CNNs (Chollet, 2022).

There are a number of features to ANNs in addition to the input, hidden and output layers. These are activation functions, weights, loss functions and optimizers and generally operate complementarily to enable the ANN to “learn”. Activation functions are employed in ANNs to determining how representations are transformed in a multi-dimensional space. They introduce non-linearity into the network, allowing it to learn complex patterns. Common examples include non-linear functions like rectified linear unit (ReLU), sigmoid, and softmax. Weights are numerical values that determine what signals are passed on by a neuron or a layer. Weights are adjusted based on the loss score, a value produced by the loss function (also known as an objective or cost function), which calculates the distance between the model's predictions and the actual targets. The learning process is focused on adjusting these weights to minimize the loss score. Common loss functions include mean squared error (MSE) and cross-entropy loss. These can also be accompanied by evaluation scores such as accuracy, mean absolute error (MAE), root mean squared error (RMSE) among others. To accomplish weight adjustments is the role of an optimizer which typically implements a gradient-based backpropagation algorithm on the loss function and weights. Various optimizers, such as RMSprop and Adam, employ specific variants of stochastic gradient descent (SGD). The operation of all these features is an iterative process that enables the network to refine its weights and improve its performance.

Prior studies have used NNs in climate interpolation for example Antonić *et al.* (2001) used the methods for the spatio-temporal interpolation of temperature variables for the Republic of Croatia. Rigol *et al.* (2001) used trend and spatial association in interpolation with ANNs for Yorkshire, UK. Jiang (2009) successfully used ANN to estimate monthly mean daily global solar radiation of eight cities in China. In several other studies, NNs have been reviewed in comparison with other interpolation methods. For example Attorre *et al.* (2007) used multilayer NNs along with IDW and UK. While UK outperformed the NNs, some of differences were not substantial and additionally NNs exhibited the highest sensitivity to local effects (Attorre *et al.*, 2007). In the study by Snell *et al.* (2000), ANNs outperformed traditional methods in the interpolation of air surface temperature. Pang *et al.* (2017) also performed a comparative study including ANNs; however, RF outperformed the rest. As in other interpolation methods, specialized variants are being developed for ANNs for example the generalized spatial autoregressive neural network for three-dimensional spatial interpolation (Zhan *et al.*, 2023).

In the majority of the studies, ANNs exhibit the potential to perform just as well as or even better than other interpolation methods, especially considering their flexibility with working with non-linear relationships, which is a caveat for the kriging methods (Rigol *et al.*, 2001; Sekulić *et al.*, 2020). This makes them particularly well-suited for interpolating climate grids in regions with limited observational data (Maier & Dandy, 2000) and complex terrains (Rampal *et al.*, 2022). While most traditional methods rely on geographical location and occasionally elevation as input variables, these few inputs could pose a performance problem for ANNs (Zeng *et al.*, 2020). However, current improvements in covariate information development with remote sensing techniques eliminates this hurdle (Sekulić *et al.*, 2020). Increasing input variables usually improves prediction results for ML methods (Hengl *et al.*, 2015; Workneh *et al.*, 2024). As one of the advantages RF over other interpolation methods, identification of variable importance (Pang *et al.*, 2017; Zeng *et al.*, 2020), has also been developed for NNs (Biecek, 2018; Mandler & Weigand, 2023). Additionally, recent advancement in hardware and software for artificial intelligence models can now also easily be harnessed for efficient computations with NNs (Chollet, 2022).

2.4 Future climate, GCMs and ensemble selection

Future climate is modeled against future physical and societal scenarios. These scenarios are currently studied and presented under the SSPs. SSPs are representations of how global society, demographics and economics might change over the next century. The five main SSPs are SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 with carbon dioxide (CO₂) concentrations ranging from 393 ppm to 1135 ppm emissions for the lowest (SSP1-1.9) and highest (SSP5-8.5) by 2100 respectively (Meinshausen *et al.*, 2020). The SSPs also reflect policy and social economic changes with SSP1-1.9 representing scenarios where more sustainable and clean technologies are embraced, and SSP5-8.5 represents a scenario where there is an increase in fuel-based technologies (Meinshausen *et al.*, 2020). Under high emission scenarios, Africa's average surface temperatures are projected to increase by 4-6 °C in the subtropics and 3-5 °C in the tropics (Engelbrecht *et al.*, 2015).

Future climate is projected by AOGCMs that simulate the interactions between the atmosphere, oceans, land surface, and ice, incorporating physical principles to represent climate dynamics

comprehensively. By solving equations that govern fluid motion, heat transfer, and various earth system processes, AOGCMs can predict changes in temperature, precipitation, and other climatic variables under different greenhouse gas emission scenarios (WMO, 2022). The accuracy of these models is validated through their ability to reproduce past climate patterns observed from historical data (Randall *et al.*, 2007). The majority of models are developed and reviewed within coupled inter-comparison model projects (CMIP) that started almost 30 years ago (PCMDI, 2022). CMIPs serve to provide multi-model output in a standardized format, making it accessible for analysis by the broader climate research community and other users (Eyring *et al.*, 2016). The current phase of the Coupled Model Intercomparison Project (CMIP6) has over 120 (AOGCMs) that have various representations of past, present and future climate change (Eyring *et al.*, 2016; PCMDI, 2024). With the available models, SSPs and climate variables there are tens of thousands of possible model simulations of projected climate change (Eyring *et al.*, 2016). The sixth Assessment Report (AR6) on Climate Change has already used over 50 GCMs and Earth System Models (ESMs) of CMIP6 to provide a broad context on climate change (IPCC, 2021; Masson-Delmotte *et al.*, 2021).

There is a consensus that a meticulous selection of AOGCMs is important for climate studies at regional scales (e.g., downscaling) as suitable for the end use (Barsugli *et al.*, 2013; Karmalkar *et al.*, 2019; Mahony *et al.*, 2022). Additionally a selection from the wide range of data alleviates the user from massive computing and data storage costs (Karmalkar *et al.*, 2019). However, it is crucial to select models that perform well and adequately represent uncertainty in future projections (Karmalkar *et al.*, 2019). This is a contentious process because AOGCM credibility varies significantly depending on the metric used, the spatiotemporal scale, the season, and the region (Masson & Knutti, 2011). Further, selecting a handful of models can potentially under-represent the uncertainty in future projections (Weigel *et al.*, 2010). Therefore selecting a representative ensemble of models involves consideration of the appropriate number and candidates. According to Mote *et al.* (2016), a minimum of 8-10 models in an ensemble are suggested to achieve a robust estimate of the mean value of some global and regional quantities. Along with that, Mahony *et al.* (2022), state that even 3-8 models may be adequate for computationally intensive models provided that they have multiple simulations for the historical period to improve the representation of uncertainty, have low bias, and are of moderate to high spatial resolution.

Various methods have been developed for selecting models in climate model ensembles, each aiming to balance the need for comprehensive uncertainty representation with the constraints on the number of models. One common approach involves performance-based selection, where models are evaluated based on their ability to simulate historical climate conditions accurately. This is often done using metrics such as biases, mean squared errors (MSEs), and other statistical measures (Taylor, 2001; Gleckler *et al.*, 2008; Minaei *et al.*, 2022). Another widely used method is clustering, which groups models based on similarities in their output to ensure that the selected subset captures the diversity of the full ensemble. Techniques such as k-means clustering and hierarchical clustering are frequently employed for this purpose (Masson & Knutti, 2011; Houle *et al.*, 2012; Mendlik & Gobiet, 2016; Wilcke & Barring, 2016; Lee & Kim, 2017). Additionally, process-based evaluations focus on selecting models that correctly simulate key climate processes and large-scale circulation patterns, ensuring physical consistency across variables (Overland *et al.*, 2014; McSweeney *et al.*, 2015). Another innovative approach is the use of algorithms like the Katsavounidis-Kuo-Zhang (KKZ) algorithm (Katsavounidis *et al.*, 1994) described by (Cannon, 2015) which selects models in a way that comprehensively covers the multivariate space of climate extremes indices, enhancing the representation of variability. The KKZ has been used for model selection by (Mahony *et al.*, 2022) for North America, (Seo *et al.*, 2019) for South Korea and (Sung *et al.*, 2019) for North Korea.

3. Data and Methods

3.1 Study Area

Africa lies between latitude 37°N and 35°S and longitude 17°W and 51°E covering approximately 30 million square kilometers. For this study I used IPCC's climate reference regions that are used in CMIP6 reporting (Iturbide *et al.*, 2020). These delineations are approximately guided by the Köppen–Geiger climate classifications, which characterizes mean temperature and precipitation (Rubel & Kottek, 2010). However, the IPCC climate reference regions do not track these climate classifications directly, but rather provide coarse polygons intended for consistent summary and description of model results from CMIP6 (Iturbide *et al.*, 2020). This research covers 11 IPCC reference regions, including the Mediterranean (MED), Sahara (SAH), Western-Africa (WAF), Central-Africa (CAF), North-Eastern-Africa (NEAF), South-Eastern-Africa (SEAF), West-Southern-Africa (WSAF), East-Southern Africa (ESAF), Madagascar (MDG) and the Arabian-Peninsula (ARP) and West central Asia (WCA) (Fig. 1).

These regions cover a wide range of climates, including tropical wet climate, tropical monsoon, tropical dry, semi-arid, desert, subtropical highland, temperate and Mediterranean climates.

Africa's climate variability is more influenced by precipitation than by temperature difference, the latter being usually consistently high all year round for majority of the continent (Hoerling *et al.*, 2006; Nicholson, 2013). Africa's climate is influenced by the position of its land mass crossed by the Equator almost halfway, by the Tropic of cancer in the north, and by the Tropic of Capricorn in the south. Areas between the Tropics experience intense solar heating that results in the high temperatures and precipitation associated with the inter-tropical convergence zone (ITCZ) or the tropical rain belt (Xie & Arkin, 1997; Nicholson, 2013). The ITCZ is characterized by circulations and convergence of warm and moist air masses that result in convectational rainfall. The ITCZ shifts towards the north in the boreal spring (March to June) and, southwards in the austral summer (September to December), influencing the number and the timing of rainy seasons in the different regions (Nicholson, 2013). Another climatic factor is the movement of global wind patterns such as the ocean currents, trade winds, westerlies and the monsoons. Ocean currents such as the Atlantic ocean's Canary and Benguela currents in western Africa, and the Indian Ocean Agulhas currents in the East influence coastal temperatures, humidity and

precipitation patterns that affect regional climates (Hoerling *et al.*, 2006). Topography, especially mountains, influence local climate by inducing precipitation through orographic lift causing higher rainfall in mountainous regions, as well as temperate and cold climate conditions, primarily in the Ethiopian highlands and rift valley ranges in SEAF and ESAF regions.

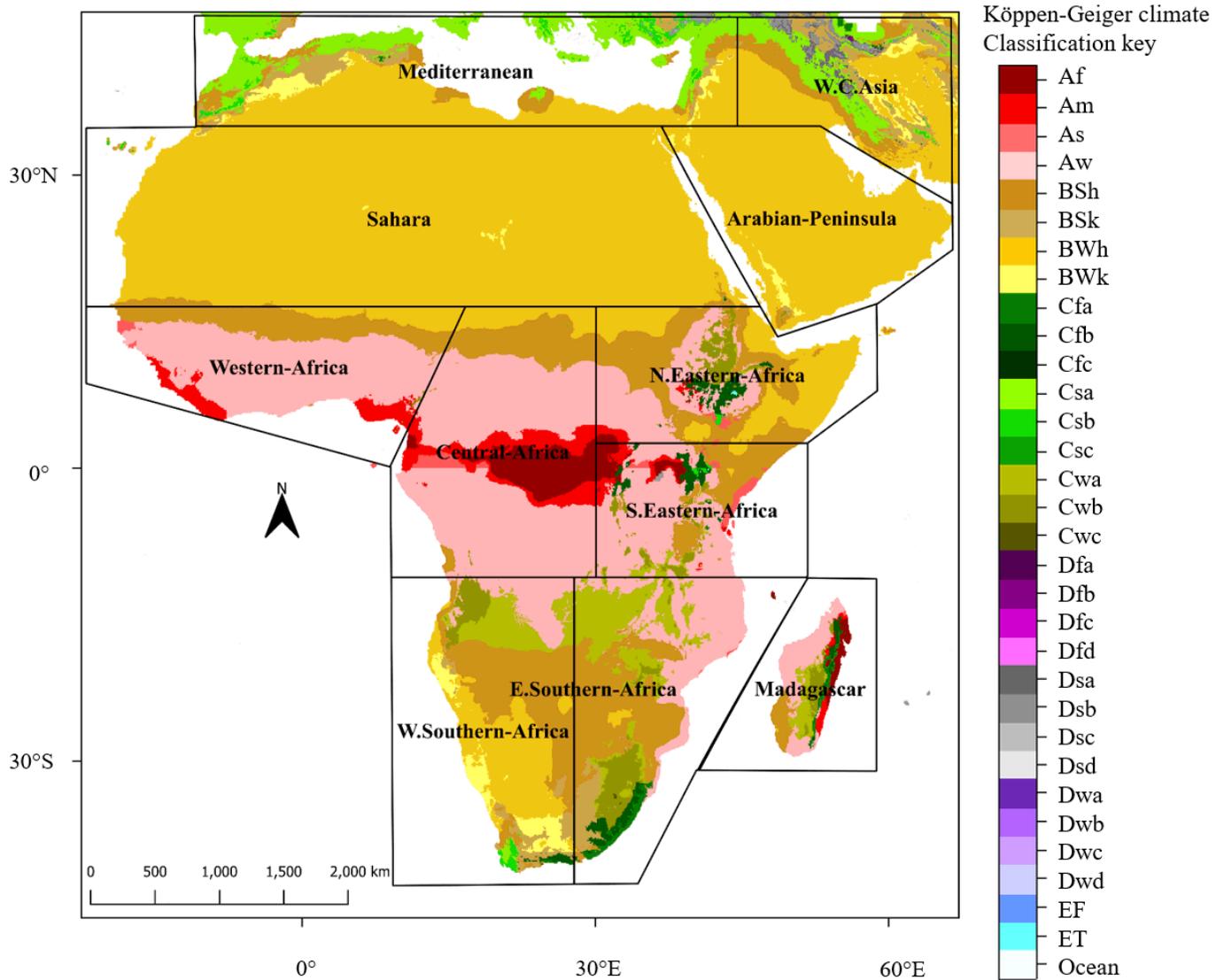


Figure 1: Africa’s IPCC climate reference regions and Köppen-Geiger climate classification (Main Climates: A – equatorial, B – arid, C – warm temperate, D – snow, E – polar; Precipitation: W – desert, S – steppe, f – fully humid, s – summer dry, w – winter dry, m – monsoonal; Temperature: h – hot arid, k – cold arid, a – hot summer, b – warm summer, c – cool summer, d – extremely continental, F – polar frost, T – polar tundra (open access data for this map was obtained from Kottek *et al.*, 2006; Iturbide *et al.*, 2020)).

3.2 Weather station data compilation

I compiled temperature weather station data from seven public databases (Table 1) that were merged and then cleared of duplicates, using similar quality control procedures as a previous compilation effort for weather station data for precipitation by Castellanos-Acuna and Hamann (2020). I compiled data from the Climate Research Unit (CRU) at the University of East Anglia, the Food and Agricultural Organization of the United Nations (FAO) Agro-Met, the World Meteorological Organization (WMO), the Global Historical Climatology Network (GHCN) managed by the National Center for Environmental Information (NCEI), the National Oceanic & Atmospheric Administration (NOAA), and the European Climate Assessment (ECA) by the Royal Netherlands Meteorological Institute (KNMI). For precipitation weather stations, I utilized the global precipitation database by Castellanos-Acuna and Hamann (2020).

Table 1. Databases included in this study detailing their temporal extent, temporal resolution and the number of stations obtained for Africa. The number of stations in parentheses are the stations that were selected after quality control and duplicate removal.

Database	Temporal extent	Temporal resolution	Number of stations	Reference
Climate Research Unit Time Series (CRUTS), Version 3	1849-2023	Monthly time series	1522 (572)	(Harris <i>et al.</i> , 2014)
Global Historic Climate Network Monthly (GHCN-M) , Version 3	1878-2017	Monthly time series	864 (176)	(Lawrimore <i>et al.</i> , 2011)
Global Historic Climate Network Daily (GHCN-D), Version 3	1900-2021	Daily time series	878 (344)	Menne <i>et al.</i> (2012)
World-wide Agroclimatic Data of FAO (FAOCLIM), Version 2	1902-1998	Monthly time series	846 (96)	(FAO, 2001)
World Meteorological Organization (WMO)	1961-1992	Monthly time series	431 (256)	(WMO, 1996)
European Climate Assessment Dataset (ECA)	1892-2018	Daily & Monthly time series	223 (13)	(Tank <i>et al.</i> , 2002)
National Oceanic & Atmospheric Administration (NOAA)	1949-2015	Monthly time series	131	(NOAA, 2018)
Global monthly weather station for precipitation	1901-2010	Monthly time series	4510	(Castellanos-Acuna & Hamann, 2020)

The consolidated database had a temporal coverage from the 1800s to 2023 with the most extensive coverage for 1961-1990 period. The dense spatial coverage during the 1961 to 1991 time period makes this climate normal period the most suitable for the development of baseline reference climate grids (New *et al.*, 1999; Menne *et al.*, 2012). Additionally, the 1961-1990

period precedes large anthropogenic climate change effects that started in the 1990s after a period of cooling induced by industrial sulfate emissions from the 1950s to 1980s, also referred to as global dimming due to post-war industrialization (Tett *et al.*, 1999; Lawrimore *et al.*, 2011).

After a master database of station records was compiled from all seven databases, the data cleaning and quality control checks included a comparison of the recorded station elevation against the digital elevation model (DEM) at 1km resolution (Gesch *et al.*, 1999). Missing values in the recorded station elevation usually denoted by -9999, -999, -99, 9999 were replaced by the DEM value. I also replaced records with a value of zero with DEM values, since records of zero were either used to indicate missing values or areas of zero elevation. The DEM value near the coast or other was likely more reliable than the recorded value of zero. In cases where the difference between recorded elevation and DEM elevation for the recorded location exceeded 250 m, I carried out manual inspection on these stations for potential errors in location.

Generally, stations with large elevation errors in mountainous regions were retained, and stations with large elevation errors in areas of flat topography were removed due to probable location errors. These checks were carried out due to the importance of the elevation in climate modeling and interpolation (DeGaetano & Belcher, 2007).

Next, I ranked all stations based on the length of the station records and their overlap with the 1961-1990 normal period. Tier 1 stations had at least 27 years of records (90% of a normal period) for the 1961-1990 period, tier 2 stations had at least 27 years of records for the 1951-1990 period. Tier 3 stations also had at least 27 years records for the 1951-2000 period. Finally, tier 4 stations had at least 15 years of records at any time from 1901 to 2020. Missing values for the 1961-1990 period were then calculated using the anomaly (change factor) approach relative to CRU-TS time series data as in Castellanos-Acuna and Hamann (2020), where the average difference between available station data and corresponding CRU-TS interpolated data was added to the CRU-TS estimates for years of missing station data. Similarly, if station records reported only average temperature, the average minimum and average maximum monthly temperatures were inferred from the interpolated diurnal range of CRU-TS time series data. Lastly the 1961-1990 normal estimate was obtained by averaging observed and estimated monthly climate values for this period.

Finally, duplicate stations in the combined databases were removed. First, I crosschecked station IDs from different databases using the enhanced master station history report (EMSHR) and its vector layer (Vose *et al.*, 2011; ESRI, 2019). This identified duplicate stations even if they had different IDs as allocated by their parent databases. Duplicates at this stage were removed by retaining the station with the higher quality tier. In a second pass of duplicate removal by location, only the highest-tier station per rounded 0.1 decimal degree (approximately 10 km grid cell size) and within the same 250 m elevation interval was retained. This ensured removal of lower quality station data where better records were available in close vicinity. I assigned a unique locator ID to the remaining stations using the database name, its parent-database ID, station name and its latitude values.

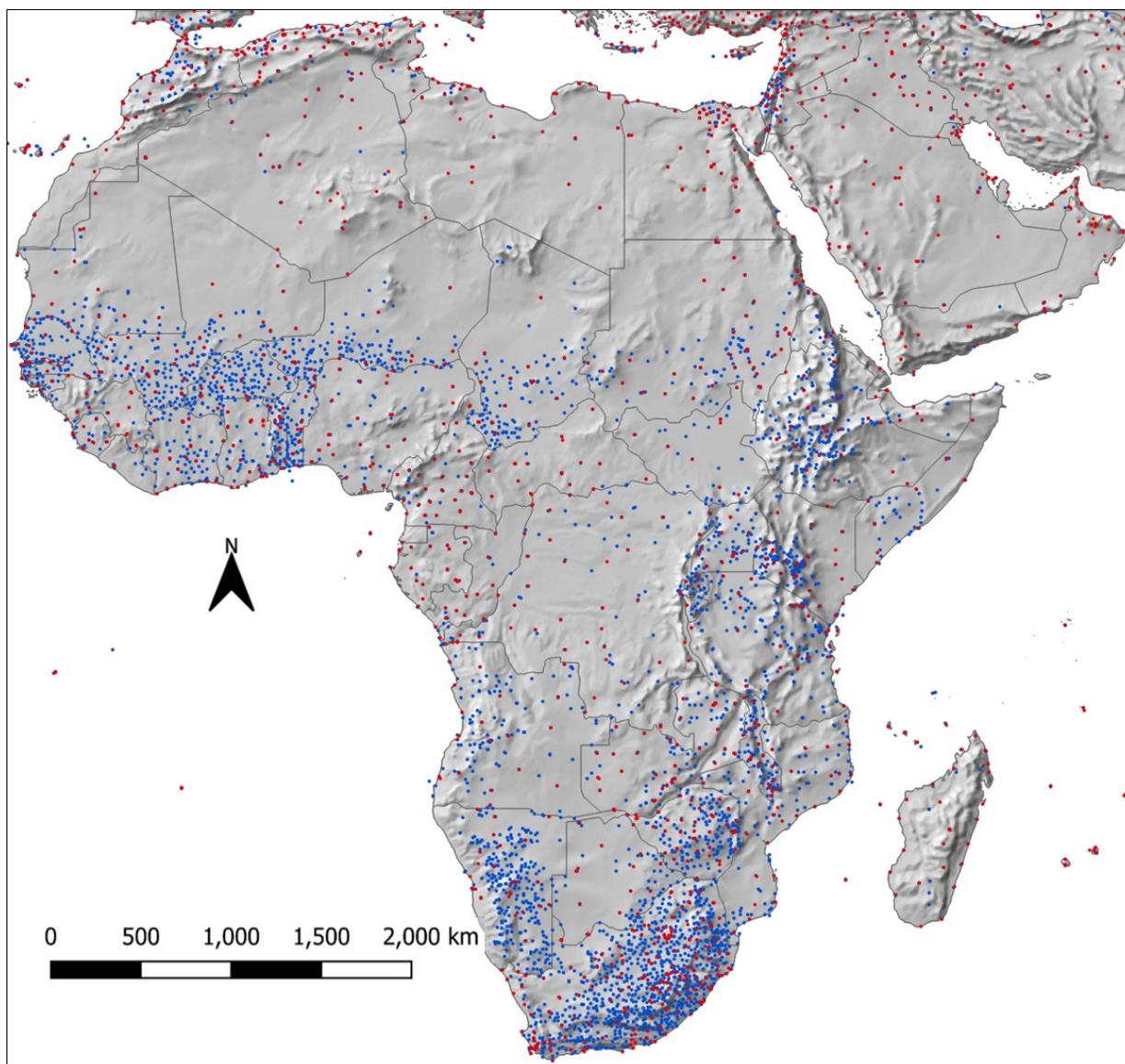


Figure 2: Distribution of 4625 weather stations compiled for the database. Blue stations have records for only precipitation measurements and red stations have records for both precipitation and temperature measurements for the 1961-1990 period.

After this quality control process, I retained 1588 stations with temperature records and 4510 stations with precipitation records. Overall, there were 4625 stations in the database with the majority only having precipitation records and some having both temperature and precipitation records. Only few (less than 20) had only temperature records. This is because historically, many agricultural and hydrological studies have prioritized collection of precipitation data over temperature data. Rainfall information is crucial for understanding water resources, irrigation

needs, and flood risks, which has led to more widespread installation of rain gauges. The spatial distribution of stations varied by region with higher station density corresponding to areas of higher population such as South Africa, West Africa and parts of Eastern Africa (Fig.2).

3.3 Thin-plate spline interpolation

A first approximation of monthly climate grids was generated with thin-plate spline interpolation, performed using the *fastTps()* function of the *fields* package (Nychka *et al.*, 2021) for the R programming environment (R). This method is computationally as well as memory intensive, and for efficiency, the interpolation was carried out on weather station data binned by three degree latitude and longitude grid cells and 250 m elevation classes (i.e. a 3D grid), with the predictor variables being the average latitude, longitude, and elevation per 3D grid cell. To further enhance computational efficiency of the interpolation, the *aRange* parameter was set to 2,000 km. This parameter controls the range of influence for the thin-plate splines. A high *aRange* value, as used here, allows the model to capture broader spatial trends, while a lower value would make the model more sensitive to local variations, if supported by a corresponding high density of weather station data (which was not available due to prior binning). A parameter enabling the function to account for the spherical geometry of the Earth's surface was set to true, ensuring that distances and spatial relationships are correctly represented in the interpolation process. The R code for the TPS model used is provided in Appendix A.

Although the thin-plate spline method would generally allow more variables (covariates) to be specified in the interpolation model, this approach is computationally too demanding because it would need be carried out on original weather station locations to capture interactions among multiple factors, simultaneously raising model complexity and the size of the dataset to be fitted. Instead I used artificial neural networks, which have practically no limitations regarding model complexity and database size, to fine tune the initial interpolated climate grids with the help of covariates reflecting topographic and geographic information (such as aspect, slope, distance to coast and lakes) in combination with monthly wind direction and strength obtained from general circulation models.

3.4 Predictor variables (features)

Predictor variables, also referred to as features by the machine learning community, were generated with *Spatial Analyst Tools* in ArcGIS (ESRI, 2008) from a 2.5 arcminute resolution DEM (the target resolution of climate grids), which was generated from GTOPO30 data (USGS, 1996; Gesch *et al.*, 1999). As potentially useful covariates besides latitude, longitude and elevation, a north-south directional hill shade, a topographic position index (TPI), which is a numerical index that describes ridges (high values), valleys (low values) and flat areas (intermediate values) were calculated. A compound topographic index (CTI) is similar to TPI, but identifies valleys and ridges with a hydrological approach where areas of convergence receive high values. The predictor variables Elevation, CTI and TPI were first transformed to be approximately normally distributed (log transformation with an appropriate constant), while the north-south directional hill shade was subjected to a bi-directional log transformation, separately for negative north-facing values and positive south-facing values to mitigate long-tailed variable distributions. Subsequently all variable values were scaled from 0 to 1. This dual transformation procedure generally improves stability and equal sensitivity to putative predictor variables during neural network training.

Further, topographic variables weighted by wind direction were generated in a two-step process, that calculated directional exposure of geographic features, which were then scaled by average monthly wind direction and strength for the 1961-1990 period obtained from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) (Gelaro *et al.*, 2017). Westward and southward exposure of mountains was approximated with hill shades (45° angle with a 180 or 270° azimuth). Directional lake and coastal influences were generated with an equivalent “topography”, derived from a 2.5 arcminute grid representing lakes or oceans with a value of 1, versus land represented by a value of 0. The grid was repeatedly subjected to a 3×3 low pass filter to the desired range of putative lake or ocean effects (approximately 10, 50, 100 and 500 km). Directional information of distance to lake shores and coast lines were generated with hill shades as described for mountains above (180 & 270° azimuth). Both topographic and distance to waterbody hill-shades were subsequently scaled from +1 (e.g., maximum westward mountain exposure, or minimum westward distance to lake or coastline) to 0 (flat topography or beyond maximum distance to waterbody), to -1 (equivalent in opposite direction) and multiplied by MERRA-2 monthly wind direction and strength, also provided in north-south and east-west direction. To generate a single exposure layer from two directional layers, the geometric mean of

east-west and north-south directional effects (geometric mean required to avoid an over-estimate of where north-south and east-west exposures overlap) was used. The resulting grids (12 exposure layers representing each month of the year) were then again re-scaled from 0 to 1 for machine learning.

As an additional step, all covariates were treated to low-pass filters of 3x3, 5x5, 7x7, 9x9 and 15x15 grid cells of the 2.5 arcminute target resolution. This is because topography and atmospheric circulations interact at different scales. For example, rain induced by orographic lift in mountainous regions takes place at the height of cloud layers and therefore does not closely track minor topographic variation at the ground level. Because the optimal scale is unknown, a range of scales for predictor variables that were evaluated through neural network importance values (based on their empirical usage in neural network weights) were generated. This first pass of neural network training was used guide the final selection of predictor variables that was used for all model runs for computational efficiency and simplicity in programming (Table 2).

Table 2. Predictor variables (features) for training the neural network. The original target resolution was 2.5 arcminutes, and low-pass filters were applied to better predict larger scale climate patters driven by higher altitude air circulation patterns.

Predictor variables for machine learning	Low-pass filter versions		
<u>Base variables</u>			
Thin-plate spline interpolation of climate variable			
Latitude			
Longitude			
<u>Topographic variables</u>			
Elevation	3	7	15
Compound topographic index	5	9	
Topographic position index	3	7	
Hill shade south-north direction	7		
<u>Monthly variables weighted by wind direction and strength</u>			
Windward slope exposure	5	9	15
Leeward slope exposure	5	9	15
Windward distance to coast (max 50km)	5		
Windward distance to coast (max 500km)	15		
Windward distance to lakes (max 10km)	5		
Windward distance to lakes (max 100km)	15		

In the study, neural network predictions were not sensitive to the exact selection of variables, but variable importance values nevertheless varied for different variables in different months. Variable importance values were generated by the DALEX package in R (Biecek, 2018).

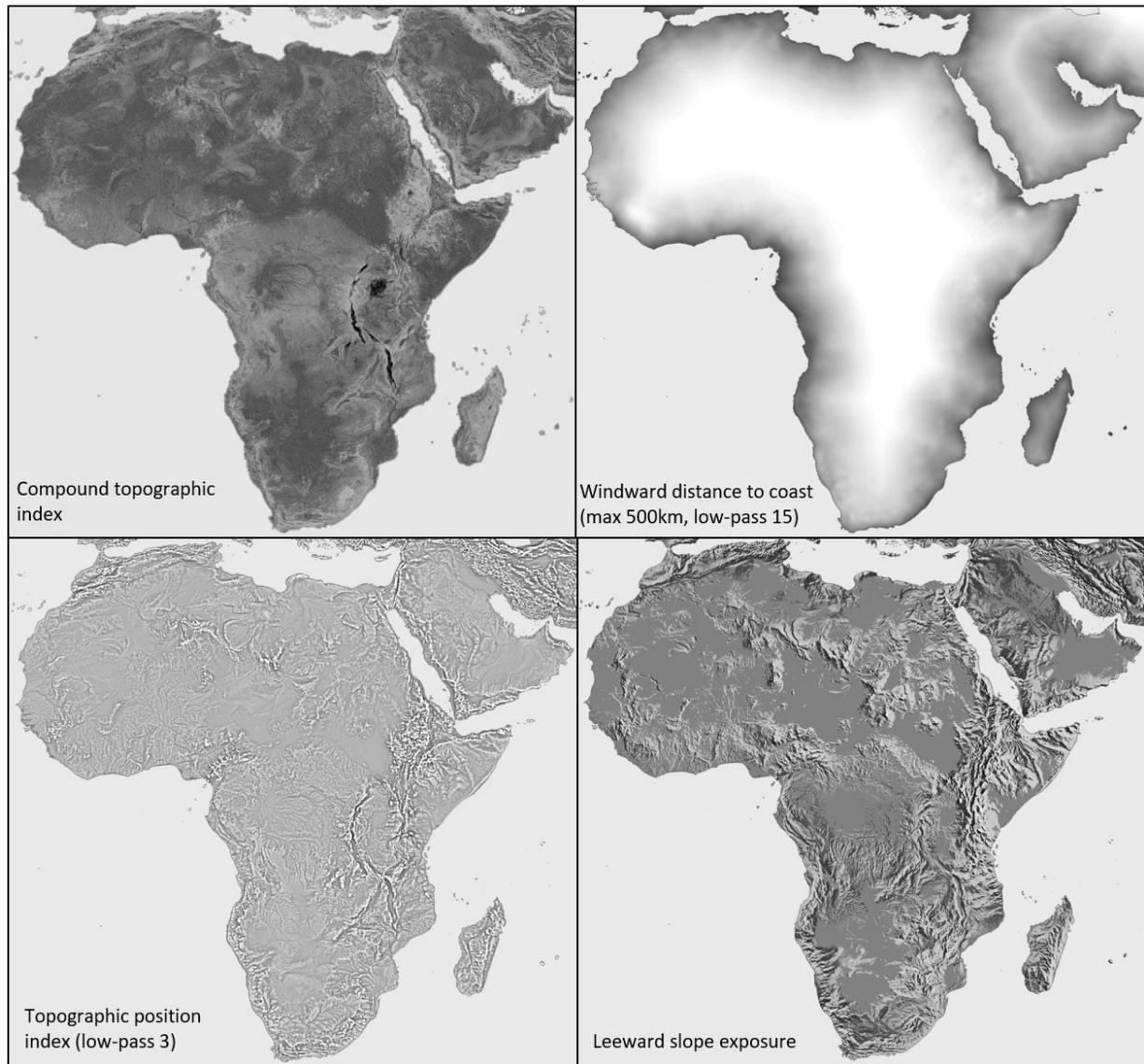


Figure 3: Example of predictor variables used in neural network fine-tuning of thin-plate spline interpolations. All putative predictor variables were subjected to transformations for normality if possible, and then scaled to values between 0 and 1 for use as covariates in neural network models.

Examples for predictor variables are shown in Fig. 3, with distance and exposure layers scaled by wind direction and strength. The rationale for these layers is that local climate patterns can be driven by the interaction of topography with atmospheric circulations.

3.5 Fine tuning with neural networks

Neural network methodology was applied to fine-tune a simplified (but computationally fast) thin-plate spline interpolation with a larger dataset of predictor variables (Table 2), including the original weather station location information as predictors, and the original climate records of stations as dependent variable. I used the Keras package for R (Allaire & Chollet, 2023) as a front end to define the neural network architecture, and the DALEX package (Biecek, 2018) to evaluate the importance of predictor variables in the neural network model. However, the computational work was programmed to be executed by Google's Tensorflow machine learning platform on an Nvidia's RTX A2000 graphics card. The network architecture is an approximately 3 million parameter model that can execute training and prediction of a single variable for a continent in about 15 minutes on any modern mid-level consumer graphics card. Note, that software compatibility is version-sensitive. As of 2024, a compatible software chain comprises Anaconda v.2022.10 with Python v.3.9.13, Tensorflow v.2.10.1, Nvidia's cuDNN v.8.1.0 library, the CUDAtoolkit v.11.2, in combination with R v4.4.x and Keras (for R) v2.13.

The neural network architecture I used was a feed forward model that was empirically optimized by varying model parameters, and observing the resulting changes in the evolution of validation statistics throughout the training period, based on an initial 80:20% random training:validation split. Model parameters that were varied included the number of hidden processing layers in the neural network (1 to 10 processing layers were tested), the number neurons per layer (8, 16, 32 ... 4096 neurons per hidden layer), and increasing, decreasing or fixed numbers of neurons from first to last processing layer. As counter-measures to over-parameterization, I further tested the inclusion of dropout layers where a proportion of neurons are re-set to a zero-activation state in specific intervals during training, as well as kernel regularization, which imposes constraints on the maximum value of the network weights in a processing layer.

As a general network architecture, that worked for all climate variables, I applied L2 kernel regularization to the first hidden layer comprising 2048 neurons, followed by a dropout layer

(rate = 20%) and seven subsequent processing layers, each half the size of the proceeding layer (1024, 512, ... 16 neurons). L2 kernel regularization is generally presented as:

$$Loss = Original\ loss + r \sum w^2$$

where r is the regularization parameter and w is the weights of the layer. A number of standard hyper-parameter choices that generally work well for feed forward network architecture proved satisfactory for our models as well: the ReLU activation function, which introduces a non-linearity into the model, a mini batch size of 32, which represents the number of stations processed at once before network weights are updated, a learning rate controlled by the Adam optimization function, and the mean squared error (MSE) as loss function. Mathematically, the ReLU function is defined as:

$$f(x) = \max(0, x)$$

where x is the input to the function and $f(x)$ is the output of the function. This simple thresholding operation helps introduce non-linearity into the network, allowing it to learn complex patterns and relationships in the data. The Adam optimization function can be defined as:

$$w_{t+1} = w_t - \frac{l}{\sqrt{\hat{v}_t} + c} \hat{m}_t$$

where w_t are the parameters at iteration t , l is the learning rate, c is a small constant added to the denominator for numerical stability, \hat{m}_t and \hat{v}_t are the bias-corrected estimates of the first mean and variance of the gradients respectively. The loss MSE which quantifies the difference between the predicted and the actual values in a dataset was defined by;

$$MSE = \frac{1}{n} \sum_{i=1}^n (o_i - p_i)^2$$

where: n is the number of samples in the dataset, o_i is the observed value of the i -th sample and p_i is the predicted value for the i -th sample.

While I tried to build a network architecture that yielded satisfactory results for all variables, the most sensitive hyper-parameter for our models were the number of epochs, which needed to be set individually for each climate variable to avoid over-parameterization (ranging from 75 to 500 passes through the station data for training). The tendency to over-parameterize was evaluated with a random 80% training and 20% validation data split in a first pass (specified as a parameter in the neural network architecture in the Keras package). The model selected for final iterations was one that performed best and had eight processing layers, one dropout layer, and was trained using 150 epochs and batch size of 32 summarized in Table 3. The R code for the model architecture and generating importance variables is provided in **Appendix B**.

Table 3. The final neural network model structure and parameter settings

Model structure		
Component	Layers	Hyper-parameters
Initial layer	Dense (nodes = 2048)	Activation = ReLU, kernel regularizer = L2 regularization
Hidden layer 1	Drop-out (rate = 20%)	
Hidden layer 2	Dense (nodes = 1024)	Activation = ReLU
Hidden layer 3	Dense (nodes = 512)	Activation = ReLU
Hidden layer 4	Dense (nodes = 265)	Activation = ReLU
Hidden layer 5	Dense (nodes = 128)	Activation = ReLU
Hidden layer 6	Dense (nodes = 64)	Activation = ReLU
Hidden layer 7	Dense (nodes = 32)	Activation = ReLU
Hidden layer 8	Dense (nodes = 16)	Activation = ReLU
Output layer	Dense (1)	
Model compilation		
Component	Type	
Optimizer	Adam	
Loss function	MSE	
Learning rate	Decaying rate (initial rate = 0.001)	
Training		
Component	Size	
Epochs	150	
Batch size	32	
Validation split	0.2 (for model development, but removed for final model)	

3.6 Checkerboard validation

While the initial model was developed with an 80% training and 20% validation data split for convenience (specified as a parameter in the neural network architecture in the Keras package for the R programming environment), this type of validation is not completely independent due to spatial autocorrelations among nearby weather stations. In other words, the samples used to train the neural network were not completely independent.

I therefore also confirmed the final network parameters using a cross-validation approach that better controls spatial autocorrelations, namely the “checkerboard” method used for the development of WorldClim (Fick & Hijmans, 2017). I used a 3° x 3° checkerboard grid to assign stations to cross-validation groups, delineated by their geographical coordinates falling within either a “black” or “white” tile. This ensured that testing data were generally distanced from training data (Fig. 4). Using stations from exclusively one group (say in white tiles) approximating to half the stations for each fold, I trained the ANN model. Subsequently, I validated the trained model on the withheld stations (in gray tiles) by using it predict the variable values. On the validation set, I calculated the mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2) between predicted estimates and observed values to evaluate the performance of the model (Li & Heap, 2008). The evaluation equations are;

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - o_i|$$

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2 \right]^{1/2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n (p_i - \bar{o})^2}$$

where n is the number of samples in the dataset, o_i is the observed value of the i -th sample, p_i is the predicted value for the i -th sample, \bar{o} is the mean of the observed values. Final climate surfaces were generated using the model that had the lowest error values in the cross-validation.

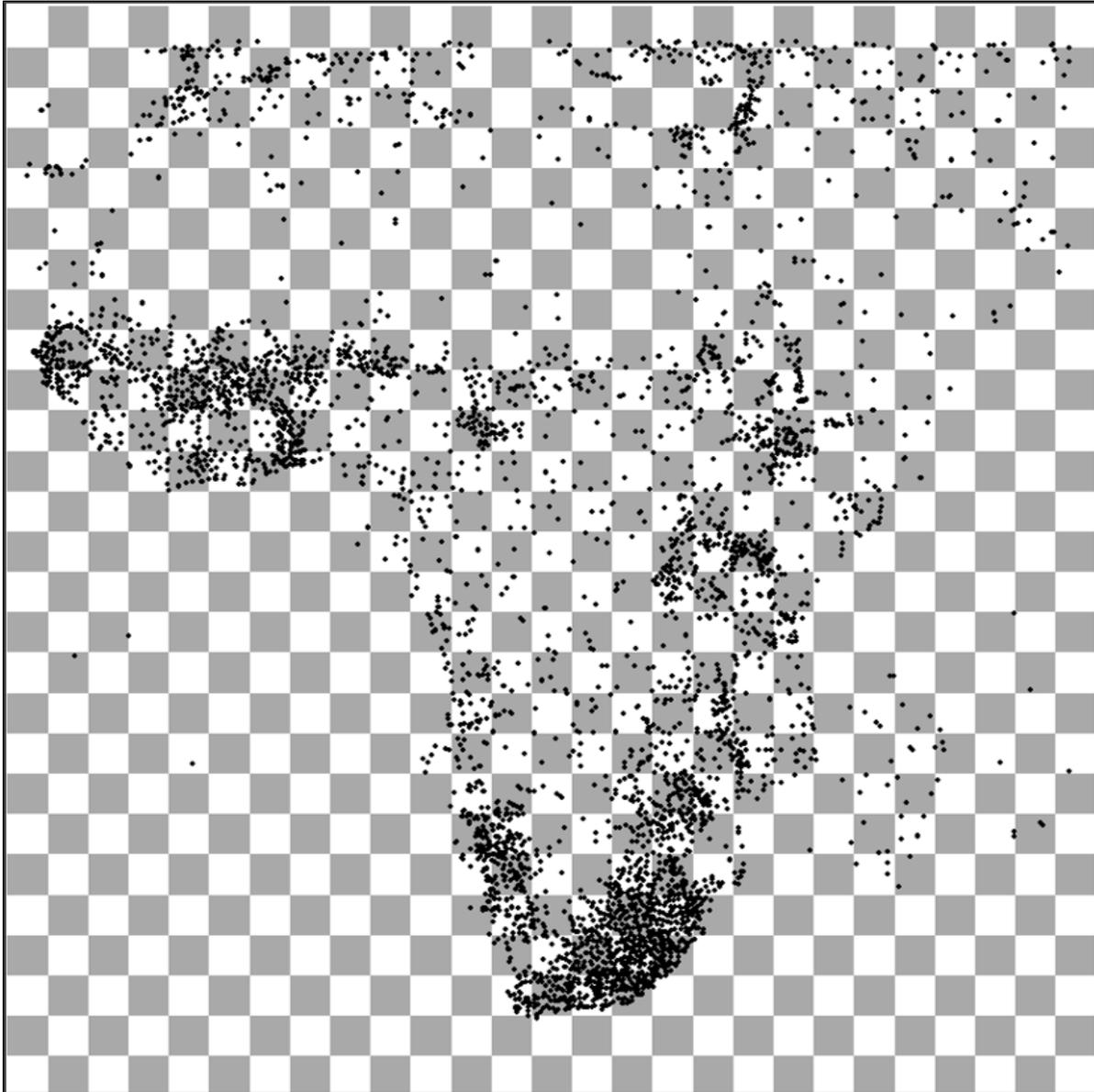


Figure 4. $3^\circ \times 3^\circ$ checkerboard grid over Africa for cross-validation groups. Black points are the weather station location falling in either the “white tile” for training or “black/gray tile” for testing.

3.7 CMIP6 models cluster analysis and regional ordered subset selection

Thirteen GCMs from CMIP6 were selected based on the six criteria used by Mahony *et al.* (2022). These filtering criteria for selecting models from the Earth System Grid Federation (ESGF) considered that models had values for both mean daily minimum temperature and mean

daily maximum temperature, a minimum of three historical runs, complete scenarios, models were not from the same institution, models were not closely related to one another, and models had no large biases (Table 4).

Table 4. Models pre-selected by Mahony *et al.* (2022) in the 13-model ensemble with their institution and citation

Model	Institution	Citation
ACCESS-ESM15	Commonwealth Scientific and Industrial Research Organization (Australia)	(Ziehn <i>et al.</i> , 2020)
BCC-CSM2	Beijing Climate Center (China)	(Wu <i>et al.</i> , 2019)
CanESM5	Canadian Centre for Climate Modelling and Analysis (Canada)	(Swart <i>et al.</i> , 2019)
CNRM-ESM2-1	CNRM (Centre National de Recherches Meteorologiques) and CERFACS (Centre Europeen de Recherche et de Formation Avancee en Calcul Scientifique) (France)	(Séférian <i>et al.</i> , 2019)
EC-Earth3	EC-Earth Consortium (European Community)	(Döscher <i>et al.</i> , 2022)
GFDL-ESM4	National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory (USA)	(Dunne <i>et al.</i> , 2020)
GISS-E2.1	Goddard Institute for Space Studies (USA)	(Kelley <i>et al.</i> , 2020)
INM-CM5.0	Institute for Numerical Mathematics (Russia)	(Volodin <i>et al.</i> , 2017)
IPSL-CM6A-LR	Institut Pierre Simon Laplace (France)	(Boucher <i>et al.</i> , 2020)
MIROC6	JAMSTEC (Japan Agency for Marine-Earth Science and Technology), AORI (Atmosphere and Ocean Research Institute), NIES (National Institute for Environmental Studies), and R-CCS (RIKEN Center for Computational Science) (Japan)	(Tatebe <i>et al.</i> , 2019)
MPI-ESM1.2-HR	Max Planck Institute for Meteorology (Germany)	(Müller <i>et al.</i> , 2018)
MRI-ESM2.0	Meteorological Research Institute (Japan)	(Yukimoto <i>et al.</i> , 2019)
UKESM1-0-LL	Met Office Hadley Centre and Natural Environment Research Council (UK)	(Sellar <i>et al.</i> , 2019)

The average Equilibrium Climate Sensitivity (ECS) across the 13-model ensemble is 3.7°C, with a variability spanning from 1.9°C to 5.6°C. This aligns with the ECS values derived from the complete CMIP6 ensemble, which also stands at 3.7°C, with a range from 1.8°C to 5.6°C (Meehl *et al.*, 2020). I aggregated the average model projections for MAT and MAP in the 13-model ensemble for the future periods 2040, 2070 and 2100 under the SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 for each of the 11 IPCC regions in Africa. Then, I performed a cluster analysis to assess and visualize the similarity among the models using six variables represented by their winter (DJF) and spring (MAM) averages for minimum monthly temperature, maximum monthly temperature, and precipitation. I selected the period of the 2050s (2041 – 2070) under

SSP2-4.5. I used *Ward.D2* hierarchical clustering algorithm with a Euclidean distance of scaled values using the *pvclust* package in R (Suzuki *et al.*, 2019) for this cluster analysis.

For the ordered model ensemble subsets, models were further excluded based on four additional criteria by Mahony *et al.* (2022). These were that models were within the very likely equilibrium climate sensitivity (ECS) of 2 and 5°C, had sufficiently high model resolution, a higher number of simulation runs and less spatial anomalous exhibition in projections. On that account, CanESM5 was excluded, for its excessive climate sensitivity and very low horizontal resolution but UKESM1-0-LL was retained despite the high ECS for one set of the subsets to represent a high-impact, low-likelihood scenarios. BCC-CSM2-MR was excluded for only having a singular simulation for each scenario and low topographic resolution, IPSL-CM6A-LR for having isolated grid cells in predicted summer temperatures and INM-CM5.0 for its low climate sensitivity (ECS 1.9°C), having only one simulation for most scenarios and its reputation as an outlier among CMIP6 models due to its under-representation of the observed 1975–2014 global temperature trend (Liang *et al.*, 2020).

To the remaining models (8 and 9 if including UKESM1-0-LL), the KKZ (Katsavounidis *et al.*, 1994) algorithm was applied as described by Cannon (2015). The KKZ selected models in a deterministic manner choosing a set of models that best represent the spread of multivariate climate change projected by the ensemble. It ordered the models with the first model closest to the ensemble centroid and the second lying furthest away from the first. To select the third and all subsequent models, the algorithm calculates the Euclidean distance of the remaining models to the previously selected models and then assigns the minimum distance to the unselected models. The model selected for the next position was the model with the maximum distance.

4. Results

4.1 Interpolated baseline grids

Grids for 36 variables, i.e., monthly precipitation (Prec), monthly minimum temperature (Tmin) and monthly maximum temperature (Tmax) for the 12 months were generated for the 1961-1990 normal period. The neural network fine-tuning was capable of capturing some complex local climate patterns that are not accounted for by standard interpolation methods. For example, for the gridded climate data for January precipitation (Prec01), the neural network was able to pick up rain shadows and precipitation due to orographic lift on the windward side of mountains in southeast Africa (Fig. 5). The inset shows that elevated terrains experience orographic rainfall, contributing to increased precipitation on the windward facing slopes (east in this case), while leeward facing slopes of the southern African mountain ranges were relatively drier due to rain shadows during the boreal winter where prevailing winds are easterly bringing moist air from the Indian ocean, contributing to high precipitation.

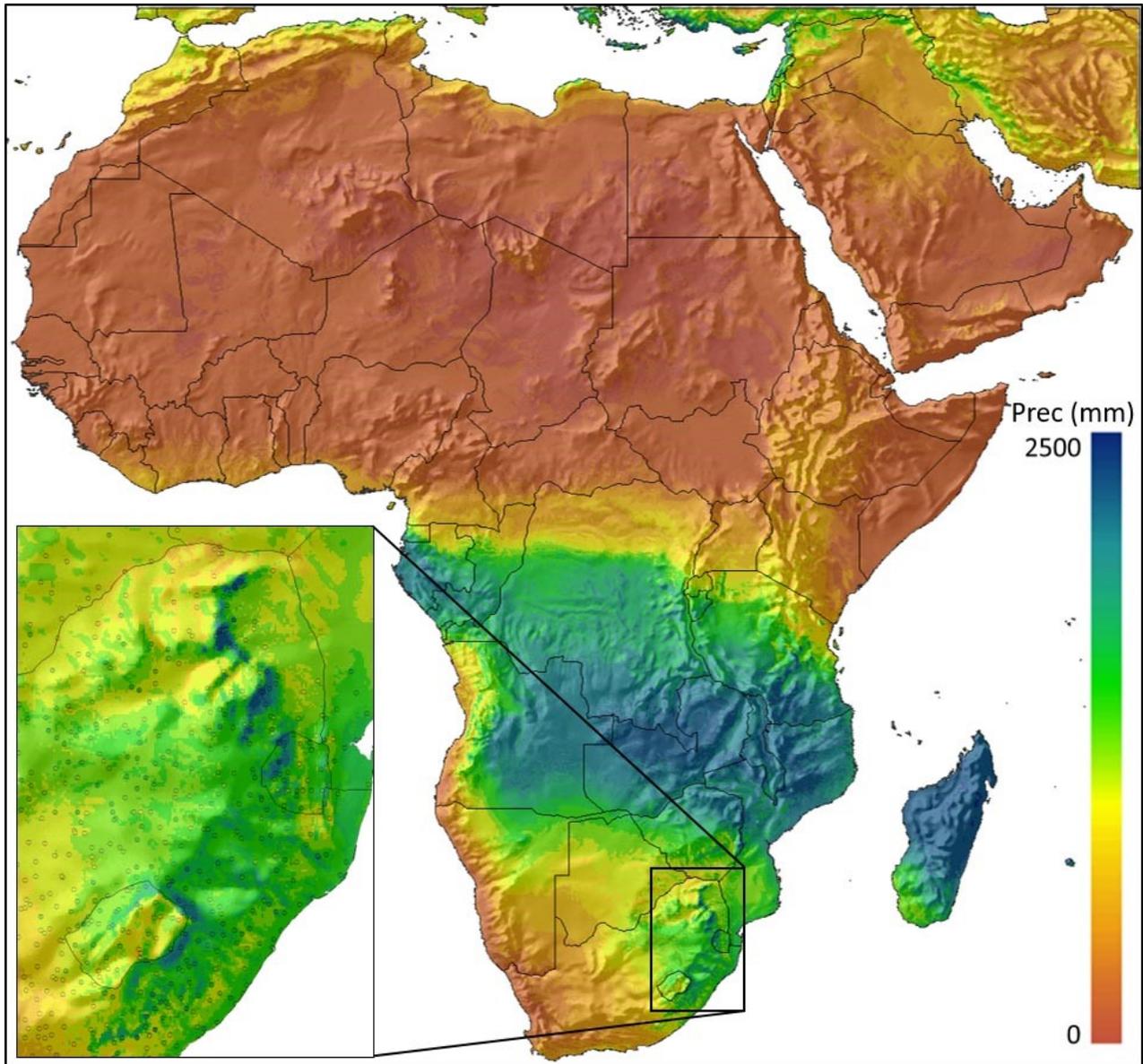


Figure 5: Example of an interpolated climate grid of January precipitation for 1961-1990. The inset highlights the influence of the micro-climatic patterns in the southeast African highlands and coast. The color of circles in the inset indicates the weather station values on the same scale.

The difference layer, between a standard thin-plate spline interpolation, and the neural network adjusted climate estimates more generally shows the fine-tuning effects (Fig 6). In general, precipitation in mountainous areas is adjusted upwards with an emphasis on windward facing slopes, whereas regions on the leeward side of mountains are adjusted downward. The inset

highlights the adjustment in the southeastern highlands where the neural network adjustments are especially prevalent.

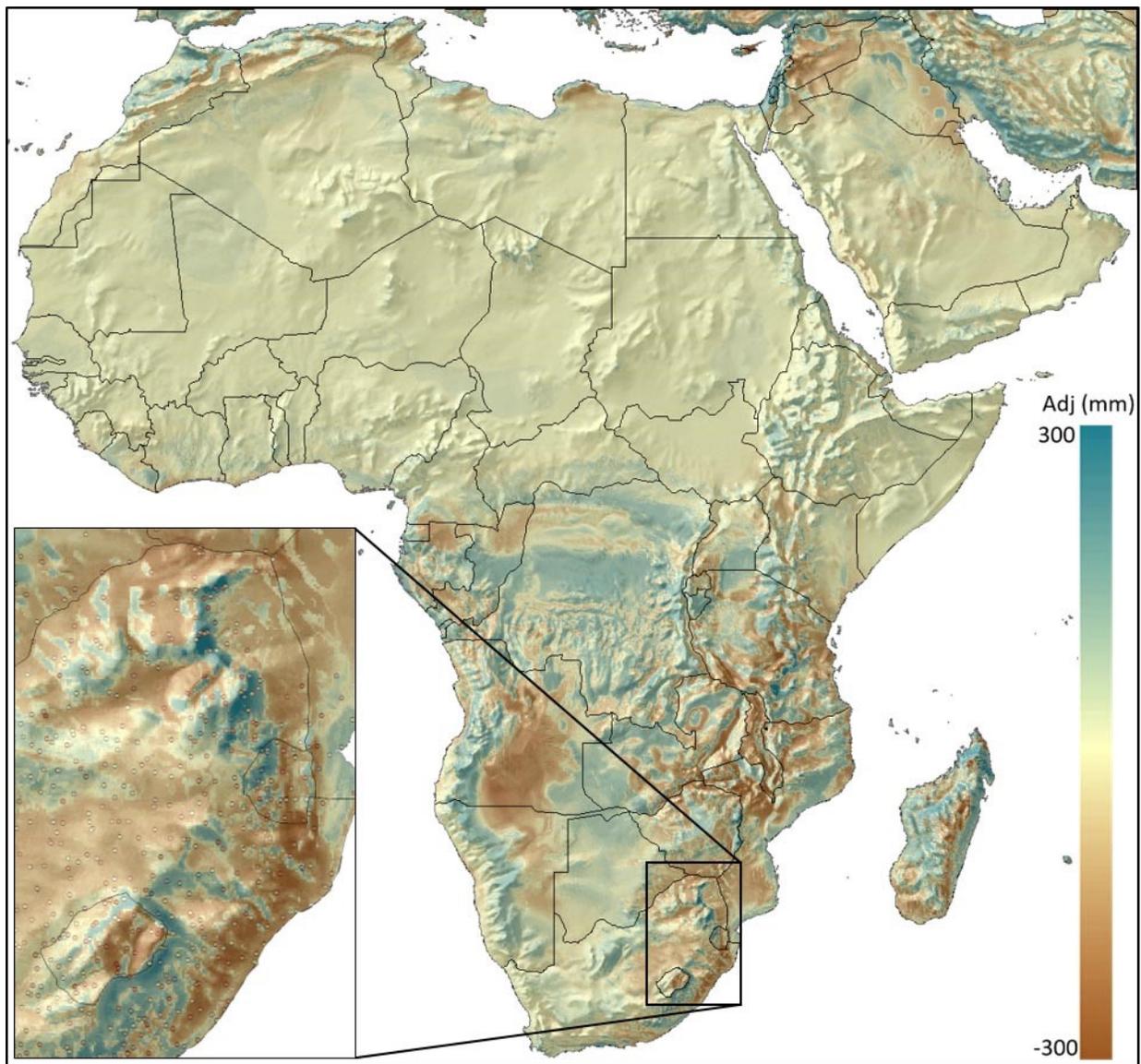


Figure 6. Adjustment to the TPS interpolated climate grid for Prec01 for 1961-1990. The inset highlights adjustment to southeast African highlands. The color of circles in the inset indicates the weather station residuals (predicted minus observed) on the same scale.

In another example for a temperature layer, minimum January temperature (T_{min01}), the inset highlights neural network adjustment for a large lake in southeastern Africa. The Lakes Tanganyika and Malawi are warmer than surrounding areas depicting water's ability to retain

and release heat more slowly than land surfaces. This also creates short-range lake effects so that surrounding areas are cooler than predicted by the thin-plate spline model. The lakes in eastern Africa such as Lake Malawi, Lake Tanganyika and, Lake Albert lie within the western arm of the African rift valley. The elevation in these locations is about 500m lower than the surrounding areas and may influence the higher temperatures in the lakes (Figs 7 and 8).

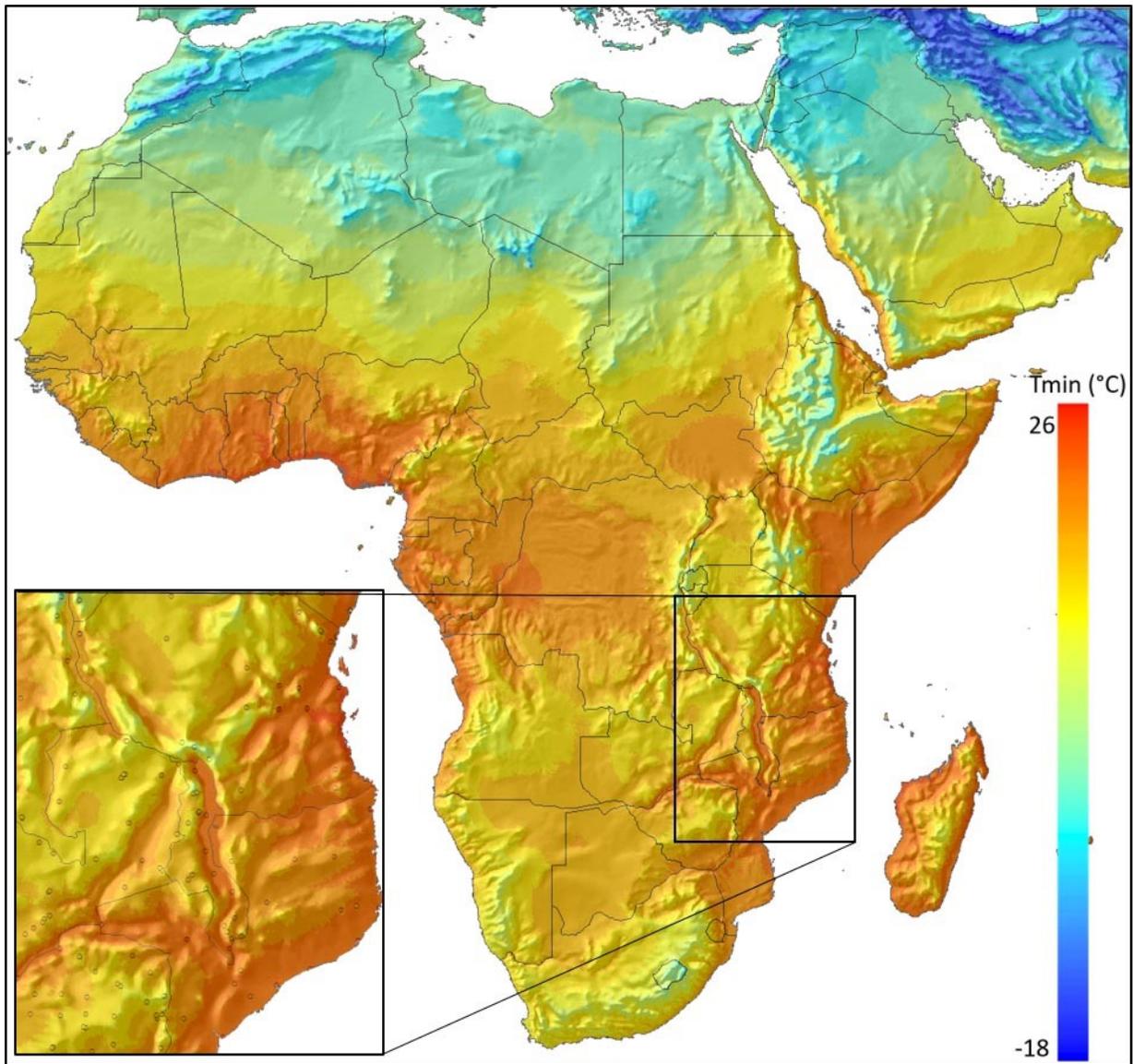


Figure 7. Example of an interpolated climate grid for January minimum monthly temperature for 1961-1990. The inset highlights the influence of the micro-climatic patterns in the southeast African highlands. The color of circles in the inset indicates the weather station values on the same scale.

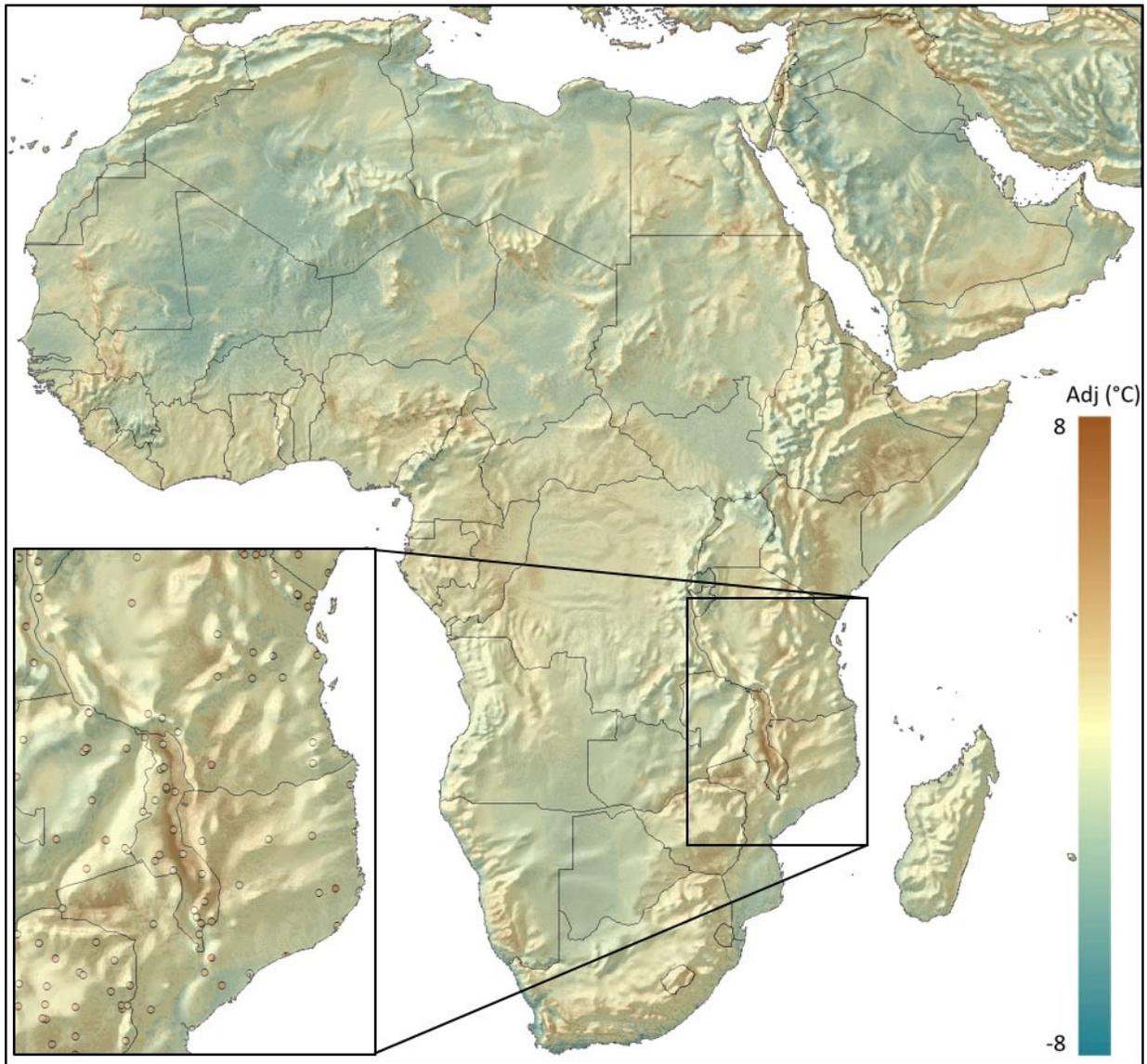


Figure 8. Interpolation adjustment by the ANN to the TPS interpolation for Tmin01. The inset is a close-up of the adjustment in the southeastern Africa around Lake Malawi. The color of circles in the inset indicates the weather station residuals (predicted minus observed) on the same scale.

In general, neural network adjustments for temperature variables were minor, indicating that the thin-plate spline interpolations already yield accurate climate estimates. Neural network importance values for features were similar for the three variables; minimum and maximum temperature and precipitation, with high ranking covariates including Latitude, Longitude,

Elevation, and windward distance to coast at short and long ranges (coast01_lp15 and coastLR01_lp15). For the precipitation model, windward slope exposure (Exp01_05), CTI and hill shade were also were picked up as important covariates (Fig 9).

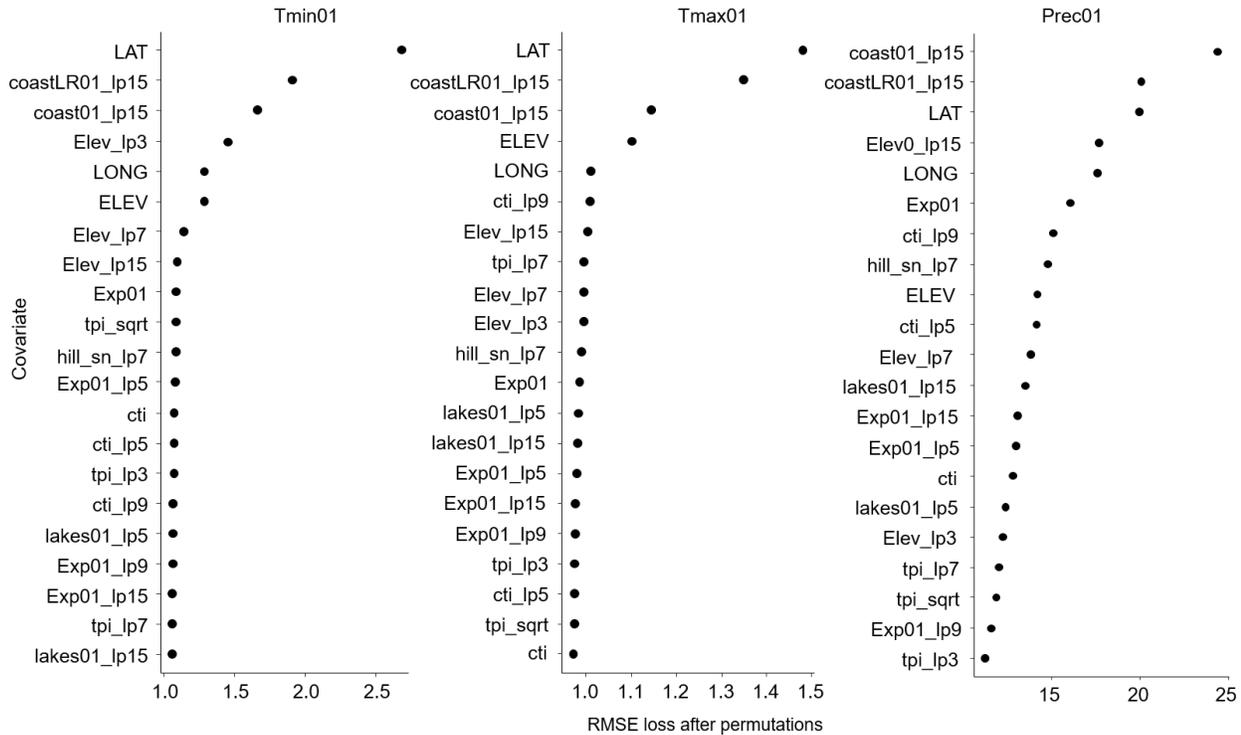


Figure 9. Feature importance for neural network interpolation for the January monthly variables average minimum temperature (Tmin01), average maximum temperature (Tmax01) and precipitation (Prec01). The candidate covariates are provided at various resolutions with low-pass filters (e.g., an “lp5” extension indicates a 5x5 grid cell average for 2.5 arcminute grid cells).

4.2 Validation of climate estimates

The accuracy of climate estimates, based on the independent “checkerboard” validation method that controls for spatial autocorrelations in weather station data varied across the variables (Tmin, Tmax and Prec) and months (Table 5). The mean absolute error values (MAE) in original units of temperature and precipitation ranged from 0.98 – 1.32 °C for Tmin and 0.96 – 1.28 °C for Tmax. For both variables, February (month 2) had the lowest MAE. June, July and August had the highest error for Tmin, Tmax. The MAE ranged from 17 to 19 mm for Prec. The JJA

months had the highest errors for Prec as well. The RMSE followed the same patterns as MAE but higher ranges reflected the presence of bias or outliers. As a third validation statistic, I calculated the variance explained (R-squared) in weather station data, which evaluates statistical precision but ignores systematic bias. R-squared values ranged from 0.92 to 0.96 for Tmin and 0.89 to 0.97 for Tmax. The values were lower for precipitation ranging from 0.73 to 0.86

Table 5. Cross-validation statistics for the ANN model. MAE is absolute mean error, R-squared is the coefficient of determination and RMSE is the root mean squared error. Tmin indicates monthly minimum temperature, Tmax is monthly maximum temperature and Prec is monthly precipitation.

Month	MAE			R-squared			RMSE		
	Tmin (°C)	Tmax (°C)	Prec (mm)	Tmin	Tmax	Prec	Tmin (°C)	Tmax (°C)	Prec (mm)
Jan	1.06	0.98	16	0.96	0.97	0.86	1.42	1.36	30
Feb	0.98	0.96	15	0.96	0.97	0.84	1.31	1.37	27
Mar	1.08	1.07	17	0.95	0.96	0.80	1.40	1.49	31
Apr	1.07	1.11	17	0.94	0.94	0.73	1.38	1.51	35
May	1.14	1.21	15	0.95	0.93	0.74	1.46	1.62	34
Jun	1.27	1.26	14	0.95	0.93	0.84	1.64	1.71	31
Jul	1.24	1.28	18	0.95	0.93	0.78	1.61	1.77	47
Aug	1.32	1.27	19	0.94	0.92	0.83	1.71	1.72	47
Sep	1.20	1.18	15	0.92	0.89	0.87	1.55	1.60	33
Oct	1.03	1.08	16	0.92	0.90	0.79	1.38	1.49	30
Nov	1.17	1.00	16	0.92	0.94	0.77	1.52	1.39	30
Dec	1.12	0.99	16	0.95	0.97	0.86	1.48	1.34	28

Once the baseline climate grids are incorporated into the Climate AF software packages, additional improvements of climate estimates are generated by downscaling of the 2.5 arcminute baseline grids to any elevation value of interest, using empirical lapse-rate based elevation adjustment for each variable, elevation and location. It should be noted that this additional evaluation was not an independent test because the test stations were used in model development for the interpolated surfaces. The MAE was used to quantify precision of estimates for the monthly temperature variables (Tmin, Tmax, Tave) and precipitation (Prec), seasonal average temperature (Tave_s) and seasonal precipitation (Prec_s) and, annual temperature (MAT) and annual precipitation (MAP) for the different IPCC climate regions of Africa (Table 6).

Table 6. Mean absolute error (MAE) between observed and interpolated surfaces for monthly temperature variables (Tmin, Tmax, Tave) and precipitation (Prec), seasonal average temperature (Tave_s) and seasonal precipitation (Prec_s) and, annual temperature (MAT) and annual precipitation (MAP) aggregated for the 11 IPCC climate regions for Africa: ARP: Arabian-Peninsula; CAF: Central-Africa; ESAF: East-Southern Africa; MDG: Madagascar MED: Mediterranean; NEAF: North-Eastern-Africa; SAH: Sahara; SEAF: South-Eastern-Africa; WAF: Western-Africa (WAF); WCA: West central Asia and; WSAF: West-Southern-Africa. AF represents the continental average for Africa.

Region	Monthly				Seasonal		Annual	
	Tmin (°C)	Tmax (°C)	Tave (°C)	Prec (%)	Tave (°C)	Prec (%)	Tave (°C)	Prec (%)
ARP	0.40	0.50	0.30	39	0.30	25	0.32	17
CAF	0.30	0.30	0.30	9	0.27	7	0.24	4
ESAF	0.30	0.40	0.20	13	0.20	9	0.18	6
MDG	0.40	0.30	0.20	20	0.17	15	0.16	11
MED	0.50	0.50	0.40	16	0.33	12	0.28	8
NEAF	0.40	0.40	0.30	15	0.27	11	0.26	7
SAH	0.50	0.50	0.40	33	0.32	21	0.30	9
SEAF	0.30	0.30	0.20	12	0.23	10	0.22	6
WAF	0.40	0.40	0.30	12	0.23	8	0.22	4
WCA	0.40	0.40	0.30	22	0.30	18	0.25	11
WSAF	0.30	0.40	0.30	18	0.27	14	0.28	8
AF	0.38	0.40	0.29	19	0.26	14	0.25	8

4.3 Selection of representative CMIP6 models

To provide guidance for selecting multi-model ensemble subsets of CMIP6 models for Africa's IPCC regions, I first calculated their different regional projections (Table 7). The 13 model ensemble had an average continental increase (MAT) ranging from 1.3 °C in SSP1-2.6 in the 2030s to 4.9 °C in SSP5-8.5 in 2080s. MAT increase was lowest in the 2030s and highest in the 2080s. MAT increase was also smallest in SSP1-2.6 and highest in SSP5-8.5. Average continental MAP ranged between 4% and 17% with marginal variations within the periods and scenarios. Much of the variation in the ensemble means was between regions for instance ARP, MED, SAH and WCA constantly had higher MAT increase in all the scenarios than the continental average and other regions. ESAF, MDG, MED and WSAF were projected to have reduction in MAP in all future projections.

Table 7. 13-model ensemble projected means of mean annual temperature (MAT; °C) and mean annual precipitation (MAP; %) by the 11 IPCC regions (**Figure 1**) and AF (continental average) for the future scenarios; SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5. See Table 6 for region abbreviations.

Change in MAT (%)												
Region	ssp126			ssp245			ssp370			ssp585		
	2030s	2050s	2080s									
ARP	1.6	2.3	2.2	1.6	2.7	3.5	1.6	3.1	4.8	1.7	3.5	5.7
CAF	1.1	1.6	1.7	1.1	2.0	2.7	1.1	2.4	3.9	1.1	2.5	4.4
ESAF	1.2	1.7	1.8	1.2	2.2	2.9	1.2	2.6	4.1	1.3	2.8	4.8
MDG	1.0	1.4	1.4	1.0	1.7	2.3	1.0	2.0	3.2	1.0	2.2	3.8
MED	1.6	2.2	2.2	1.6	2.6	3.3	1.6	2.9	4.5	1.7	3.3	5.3
NEAF	1.1	1.7	1.7	1.1	2.0	2.7	1.1	2.3	3.7	1.1	2.6	4.4
SAH	1.6	2.2	2.2	1.7	2.7	3.5	1.7	3.2	4.9	1.7	3.5	5.7
SEAF	1.1	1.6	1.6	1.1	2.0	2.6	1.1	2.3	3.6	1.1	2.5	4.2
WAF	1.1	1.7	1.7	1.1	2.0	2.7	1.1	2.3	3.8	1.1	2.6	4.5
WCA	1.8	2.5	2.5	1.8	2.9	3.8	1.8	3.3	5.1	1.9	3.8	6.1
WSAF	1.3	1.8	1.9	1.3	2.3	3.1	1.3	2.7	4.4	1.3	3.0	5.1
AF (Avg)	1.3	1.9	1.9	1.3	2.3	3.0	1.3	2.6	4.2	1.4	2.9	4.9

Change in MAP (°C)												
Region	ssp126			ssp245			ssp370			ssp585		
	2030s	2050s	2080s	2030s	2050s	2080s	2030s	2050s	2080s	2030s	2050s	2080s
ARP	26	34	25	26	34	43	25	43	64	30	50	73
CAF	9	10	7	10	11	12	9	14	22	10	17	25
ESAF	-6	-8	-10	-6	-10	-12	-4	-11	-15	-4	-11	-17
MDG	-6	-8	-8	-5	-8	-10	-5	-9	-13	-6	-11	-14
MED	-4	-8	-10	-3	-12	-16	-5	-13	-22	-5	-15	-27
NEAF	17	28	27	17	28	39	17	34	58	20	42	71
SAH	22	20	14	24	20	22	23	30	36	24	31	35
SEAF	5	7	6	5	8	11	5	10	18	5	12	21
WAF	10	10	5	12	9	10	9	13	20	12	15	23
WCA	8	10	4	8	7	9	7	11	19	9	13	20
WSAF	-6	-12	-13	-7	-13	-16	-5	-13	-21	-7	-16	-24
AF (Avg)	7	8	4	7	7	8	7	10	15	8	12	17

The cluster analysis and visual maps of the 13 models and ensemble mean indicated spatial similarity and difference of projected seasonal changes in minimum temperature (Tmin), maximum temperature (Tmax), and precipitation across the African continent for the winter (DJF) and spring (MAM) seasons during the period 2041-2070 (2050s) under the SSP2-4.5 scenario (Fig 10). The temperature projections (Tmin and Tmax) indicated a consistent increase across the continent, with higher changes predominantly in northern and southern Africa during

winter, and pronounced increases in Tmax, along with substantial increases in Tmin, in central and eastern Africa during spring. The precipitation projections, which are log-scaled to highlight proportional changes, show regions with both positive and negative changes. In winter, notable decreases in precipitation are observed in parts of northern Africa, while central regions see increases. During spring, similar mixed patterns are observed, with significant increases in precipitation in some areas of central and eastern Africa. All models exhibited somewhat similar patterns for precipitation albeit different magnitudes. The largest disparities between models was observed in the temperature variables. CanESM5 and UKESM1-0-LL projected the highest magnitude of temperature increase across the continent in both seasons. On the other hand, MPI-ESM1-2-HR, INM-CM5-0, MRI-ESM2-0 and MIROC6 projected declining temperature changes in both seasons especially in the central and southern parts of the continent.

In eight selected models (nine, including high sensitivity UKESM1-0-LL), the KKZ algorithm systematically selected models recommended for each region and the continent (AF) based on capturing the range of multivariate climate changes projected by the ensemble **Table 8**. The KKZ subset selection was ordered starting with the model closest to the ensemble centroid and the second model furthest away from the first model selected. The third and all subsequent positions were selected based on their distances from the previous model. For example for the WSAF, the first model closest to the ensemble centroid is EC-Earth3, the second model in the subset is UKESM1-0-LL that lies farthest away from the first model and the third is MIROC6 that has the maximum value of the assigned distances of the remaining models to the selected models and so on (Fig 11). All the regions had different orders of the ensemble subset members.

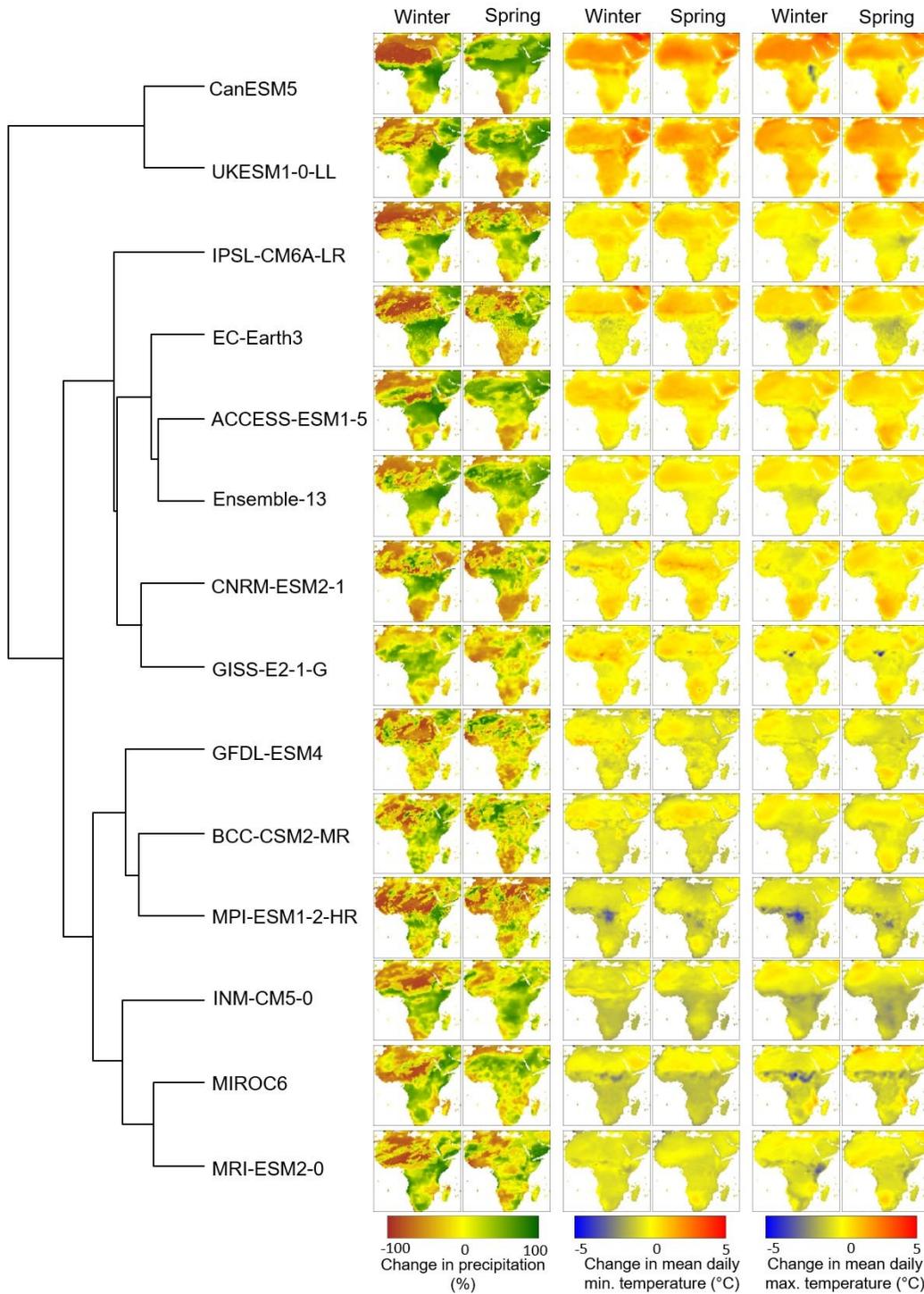


Figure 10: Models are structured by a cluster dendrogram showing spatial similarity in the projected seasonal changes for Tmin, Tmax and precipitation in winter (DJF) and spring (MAM) in the period 2041-2070 under (SSP2-4.5). The maps illustrate the visual changes across the African continent for this period. Precipitation is log-scaled to provide proportional magnitude of positive and negative changes

Table 8: Subsets of the projections with optimal representation variation in climate change projections for a given subset size according to, including additional selection criteria from Mahony *et al.* (2022). For example, a 4-model ensemble for the ARP region would include CNRM-ESM2-1, UKESM1.0-LL, EC-Earth3 and, MPI-ESM1.2-HR. Users may choose to exclude the extreme projections of UKES as well. See the Table 6 caption for region abbreviations.

Subset size	IPCC reference region											
	ARP	CAF	ESAF	MDG	MED	NEAF	SAH	SEAF	WAF	WCA	WSAF	AF
<u>Including UKESM1-0-LL</u>												
1	CNRM	GISS	EC	CNRM	CNRM	GISS	CNRM	CNRM	CNRM	CNRM	EC	CNRM
2	UKES	UKES	UKES	UKES	UKES	UKES	UKES	UKES	MPI	UKES	UKES	UKES
3	EC	MPI	MPI	MPI	EC	MIR	MPI	MPI	UKES	EC	MIR	MPI
4	MPI	EC	MIR	MIR	MPI	ACC	EC	MIR	EC	MPI	CNRM	GFDL
5	MRI	MIR	CNRM	GISS	GISS	GFDL	MRI	EC	GFDL	GFDL	GISS	ACC
6	ACC	ACC	GISS	ACC	MRI	MRI	ACC	GISS	MIR	MIR	MRI	EC
7	GISS	GFDL	MRI	MRI	ACC	EC	GISS	MRI	GISS	ACC	GFDL	MIR
8	MIR	CNRM	ACC	EC	MIR	MPI	MIR	GFDL	ACC	MRI	MPI	GISS
9	GFDL	MRI	GFDL	GFDL	GFDL	CNRM	GFDL	ACC	MRI	GISS	ACC	MRI
<u>Excluding UKESM1-0-LL</u>												
1	GISS	MRI	GFDL	MRI	CNRM	GISS	CNRM	GFDL	GISS	MRI	GFDL	GISS
2	EC	MPI	CNRM	ACC	EC	MIR	MPI	ACC	MPI	EC	CNRM	MPI
3	MRI	ACC	MPI	MIR	MPI	ACC	EC	MIR	EC	MPI	MIR	ACC
4	MPI	MIR	MIR	CNRM	GISS	GFDL	MRI	MPI	ACC	MIR	GISS	MIR
5	ACC	EC	GISS	GISS	MRI	MRI	ACC	EC	MIR	ACC	EC	MRI
6	CNRM	GFDL	EC	MPI	ACC	EC	GISS	MRI	CNRM	GFDL	MPI	EC
7	GFDL	CNRM	ACC	EC	MIR	MPI	MIR	CNRM	GFDL	CNRM	MRI	CNRM
8	MIR	GISS	MRI	GFDL	GFDL	CNRM	GFDL	GISS	MRI	GISS	ACC	GFDL

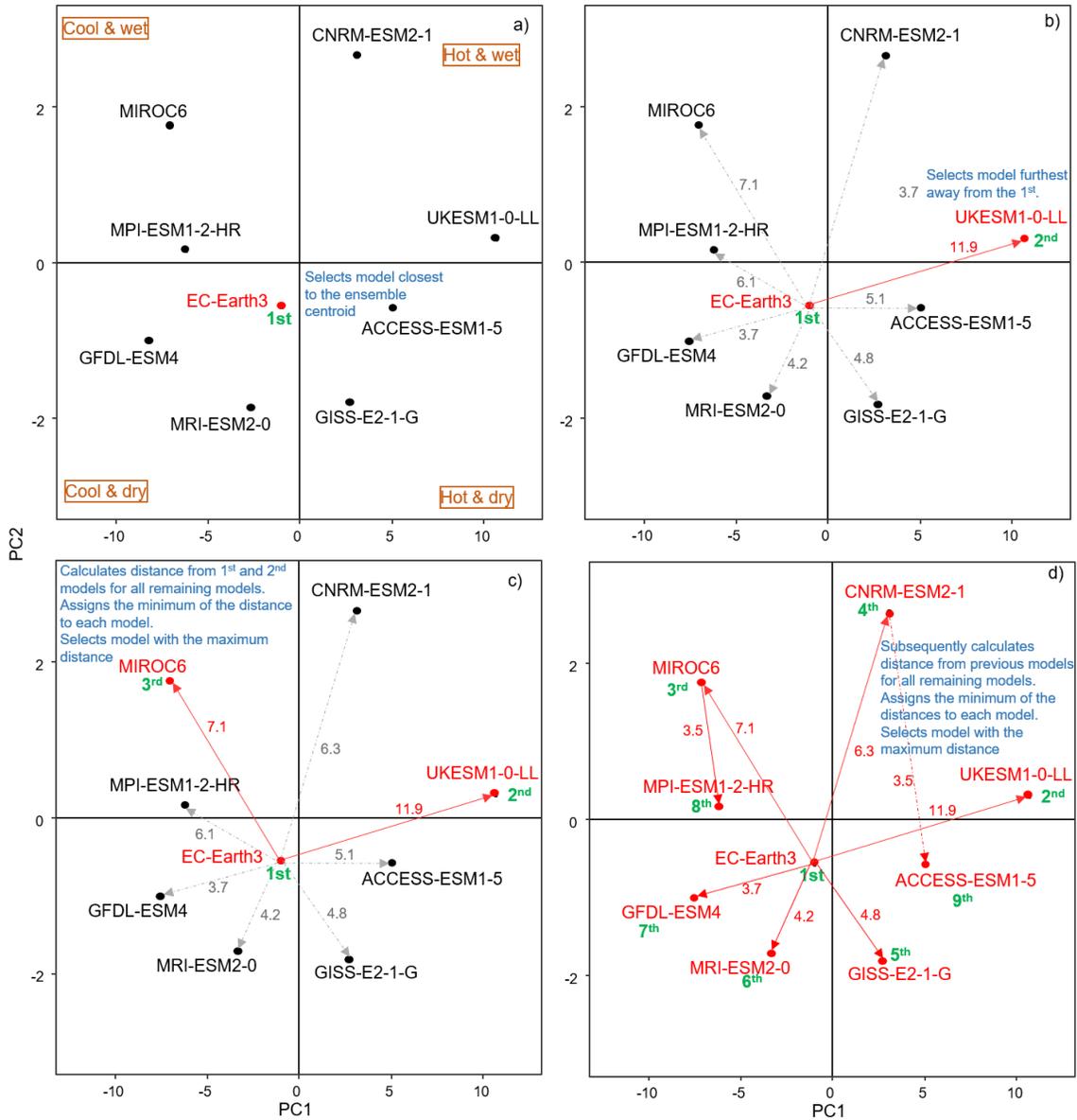


Figure 11: Illustration of the selection process by the KKZ algorithm for the region of west-southern Africa (WSAF) as an example. The first model is selected based on the ensemble centroid in multivariate space, represented here as principal component scores of climate variables (a). The four quadrants (brown labels in panel a) represent primarily temperature gradients (PC1) and precipitation gradients (PC2), with the corresponding variable loadings of the PCA provided in Table 9. The panels (b) to (d) show the sequential selection of additional scenarios to optimally represent uncertainty with given a minimum number of scenarios, where the multivariate Mahalanobis distances are shown with red numbers.

Table 9: Variable loadings of the PCA plots in Figure 11. Seasonal variables are for winter (wt; DJF), spring (sp; MAM), summer (sm; JJA) and autumn (at; SON) for the minimum monthly temperature (Tmin), maximum monthly temperature (Tmax), average monthly temperature (Tave) and precipitation (prec). Monthly maximum and minimum temperature for January to December were also used for the ordination.

Seasonal variables	Maximum monthly temperature		Minimum monthly temperature				
	PC1	PC2	PC1	PC2			
Tmin_wt	0.17	0.04	Tmax_1	0.16 0.12	Tmin_1	0.17	0.03
Tmin_sp	0.17	-0.05	Tmax_2	0.16 0.23	Tmin_2	0.17	0.06
Tmin_sm	0.15	-0.28	Tmax_3	0.16 0.20	Tmin_3	0.17	0.03
Tmin_at	0.17	-0.14	Tmax_4	0.17 0.14	Tmin_4	0.16	-0.02
Tmax_wt	0.17	0.19	Tmax_5	0.18 -0.01	Tmin_5	0.16	-0.17
Tmax_sp	0.17	0.11	Tmax_6	0.17 -0.03	Tmin_6	0.14	-0.23
Tmax_sm	0.17	-0.05	Tmax_7	0.17 -0.08	Tmin_7	0.14	-0.33
Tmax_at	0.16	0.07	Tmax_8	0.16 -0.03	Tmin_8	0.15	-0.25
Tave_wt	0.18	0.12	Tmax_9	0.16 0.08	Tmin_9	0.16	-0.18
Tave_sp	0.18	0.05	Tmax_10	0.14 0.03	Tmin_10	0.17	-0.15
Tave_sm	0.17	-0.15	Tmax_11	0.16 0.10	Tmin_11	0.16	-0.09
Tave_at	0.17	-0.04	Tmax_12	0.16 0.21	Tmin_12	0.17	0.03
Prec_wt	-0.04	-0.30					
Prec_sp	-0.13	-0.21					
Prec_sm	-0.08	-0.29					
Prec_at	0.06	-0.23					

5. Discussion

The results from neural network fine-tuning applied to thin-plate spline interpolation suggests that the approach can capture additional local climate patterns that may not be optimally represented by standard interpolation techniques alone (Fig. 12).

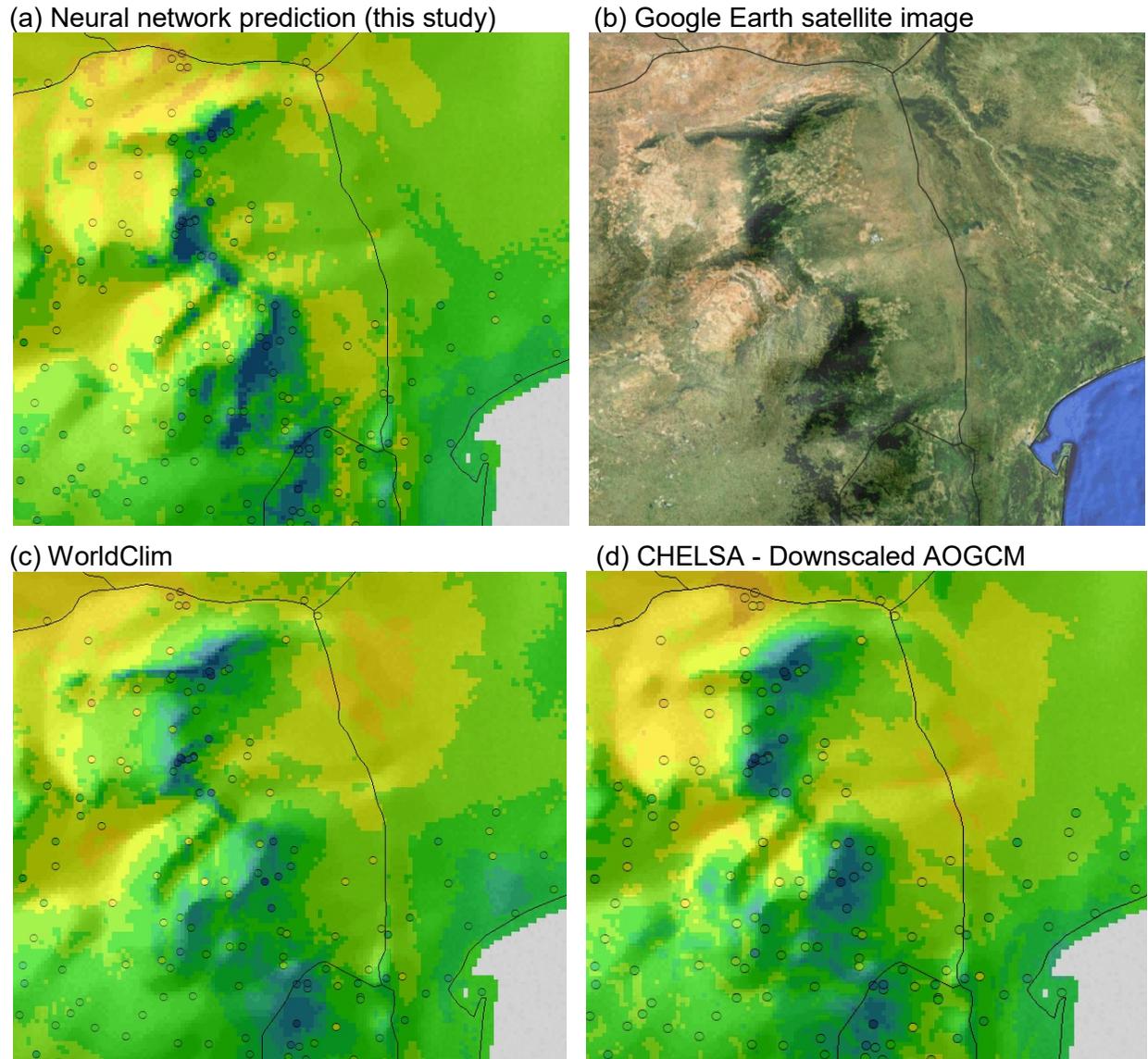


Figure 12: Comparison of interpolated climate grids for January precipitation, fine-tuned with neural networks from this study (a), with a Google Earth satellite image (b) and other popular and well-regarded data products: WorldClim (c), CHELSA (d). Weather station data is represented by open circles with fill values on the same scale as the climate interpolations.

In this example for January precipitation, the patterns of dense vegetation on windward (south east) facing slopes very closely track precipitation patterns (e.g., Fig 12 a versus 12 b). There is a distinct confinement of forests to areas where precipitation is predicted to be highest, just below the ridges, and a sharp onset of rain shadows (and lack of vegetation) beyond the ridge line on the leeward facing slope. Note that in January, southward facing slopes do not receive high precipitation, but they do in other months, accounting for small discrepancies between forest cover and high values of precipitation in this example.

The widely-used WorldClim dataset (Fick & Hijmans, 2017) models climate variables as a function of latitude, longitude and elevation, which is certainly is a good approximation of precipitation patterns, but forces elevation gradients on variables that do not always reflect weather station records (Fig 12 c). The CHELSA v2.1 dataset (Karger *et al.*, 2017), which downscales the re-analysis product ERA5 has improved representation of precipitation induced on windward facing slopes, but it does not quite capture the maximum orographic precipitation just below mountain ridges, and the rapid onset of rain shadows on the leeward slopes (Fig. 12 d).

Similarly, temperature variables are influenced by the vicinity of large water bodies, which interpolation methods cannot always predict correctly if there are not enough weather stations to allow modeling of the temperature gradient from near the shoreline to further inland. This is similar to findings by Attorre *et al.* (2007), who tested multiple interpolation methods including neural networks for a small region in Europe, and Alsafadi *et al.* (2023), who highlights the importance of using covariates to obtain more realistic representation of climate patterns in complex terrain.

The relatively minor adjustments that the neural network generated for temperature variables suggest that thin-plate spline interpolation already provides a close approximation for these variables. However, the neural network's refinements still add value by making subtle improvements primarily based on distance to water bodies, especially oceans, as well as elevation and hill shades in north-south direction. The effect of large water bodies on temperature and a buffering effect on the diurnal range of T_{min} and T_{max}, as observed in this study, are well known from other research as well (e.g., New *et al.*, 1999). The high importance

of positional covariates such as latitude, longitude, and elevation indicates that the adjustments are not universally applied to the entire continent, but that the magnitude of the effects varies with location and elevation.

The independent evaluation of the interpolated climate grids for the 1961-1990 normal period provide a realistic expectation of typical errors for monthly temperature variables (1 to 1.3°C) and monthly precipitation variables (approximately 15 mm on average), based on the mean absolute error (MAE) statistic (Table 5). The additional downscaling by the ClimateAF software through lapse-rate based elevation adjustments markedly enhances the accuracy of climate estimates as well, with remaining errors in temperature variables ranging from 0.2 to 0.5°C. Because these statistics are by necessity for a non-independent test, where all weather stations are used for the final grids to be incorporated into the ClimateAF software package, some of the improvements could be due to over-parameterization. However, we argue that this is unlikely to apply because of the following observations. The MAE statistics for precipitation in the independent checkerboard cross-validation approach (Table 5) do not change compared to the non-independent validation of the ClimateAF precipitation estimates (Table 6). This is a strong indication that the neural network tuned thin-plate spline interpolations are not an over-parameterized model. Precipitation variables do not respond to physical lapse-rate based changes in elevation, thus there are no improvements (as in temperature variables), and the validation statistics remain identical even when all station data are used for both training and validation, which would not be expected if the model over-parameterizes.

Another general observation on error statistics is that they diminish when averaged over longer time periods. Errors improve when comparing monthly, seasonal and annual variables (Table 6), which has also been observed by Mbogga *et al.* (2010) who noted that stochastic variation is evened out over longer time periods. Geographically, the precision of climate estimates varied significantly. Central and western Africa (CAF and WAF) showed the lowest percentage errors for seasonal precipitation, indicative of more stable and predictable climate patterns in these regions. Conversely, the Arabian Peninsula (ARP) and the Sahara (SAH) exhibited the highest errors, reflecting the inherent difficulty in modeling climates with extreme conditions and sparse data availability similar to findings by New *et al.* (1999), Fick and Hijmans (2017), Funk *et al.* (2015) and Novella and Thiaw (2013).

Accurate, high resolution estimation of climate variables in complex terrain has numerous applications in ecology, forestry, hydrology, agriculture and infrastructure studies. In North America and Europe, the ClimateNA and ClimateEU data packages and software front-end have been used for studies in the delineation of forest seed planning zones and assisted migration (Mekonnen *et al.*, 2019; Sáenz-Romero *et al.*, 2020; Wesselkamp *et al.*, 2024), species distributions (Jarnevich *et al.*, 2018; Illés & Móricz, 2022), climate adaptation (Lovell *et al.*, 2021), wildfires (Parks *et al.*, 2018; Cansler *et al.*, 2022; Parisien *et al.*, 2023), ecosystem services (Schirpke & Ebner, 2022; Hu *et al.*, 2023) among many other applications. Similarly the ambition of the ClimateAF database and software front end is to enable research of similar quality for the African continent. The neural network-tuned climate grids in the ClimateAF data package are qualitatively comparable to estimates from PRISM methodology (Daly 2008), which is also capable of modeling local weather patterns in complex terrain, but only available for North America and Europe (and incorporated in the ClimateEU and ClimateNA packages).

Despite the ability of these software packages to estimate climate variables at any resolution, with meaningful accuracy improvements down to 250 m resolution in mountainous terrain, it is essential to recognize the general limitations of the data underlying the climate estimates provided by the software package. Although climate data can be produced at very fine resolutions, this data is ultimately based on interpolated information from standard weather stations that are typically located in open terrain (e.g. farms, fields, airports), and while the instruments are shaded, the stations are typically installed distant to positions that are topographically shaded or shaded by vegetation. As such all climate estimates represent those open field type conditions, even with adjustments made by neural networks based on topographic positions. The topographic position estimates are only modeled at the scale native to the model (2.5 arcminutes or approximately 4 km resolution). The only other useful information that the model predicts are temperature gradients along elevation gradients, up to a useful resolution of 250 m in mountainous terrain (with a maximum of one decimal of numerical precision provided by the software).

Therefore, it should be kept in mind that while the covariates used in the baseline grid layers improve local effects of large topographic features, the models do not account at all for micro site effects caused by vegetation, small water bodies, or other small-scale physiographic features.

As noted above, errors are also larger for estimates that are not averaged over longer periods. What was observed for monthly variables for the 1961-1990 period, that seasonal or annual summaries become more accurate, further extends to multi-year averages. Climate estimates for 30-year normal periods are more precise than for decadal averages, individual historical years, or individual historical months, with the highest uncertainties associated with the latter. Further, regions in remote areas where the number of weather stations is limited such as the Sahara present larger errors and should be used cautiously. The same applies the further the predictions approach go back in time toward the beginning of the included time series data, 1901, where weather station coverage was sparse everywhere in Africa. The way that the data quality degrades in CRU-TS time series data is that in the absence of weather station records, the climate estimates gradually approach the 1961-1990 normals. This will become obvious to the user, when they plot time series data for a specific location. If no station data is available beyond a certain point back in time, the time series flat-lines for earlier years.

For future projections, the 13 AOGCM model selected by Mahony *et al.* (2022) globally conserve the climate uncertainty of the larger group of CMIP6 models. The selected models had an average equilibrium climate sensitivity (ECS) of 3.7°C similar to the ECS of the larger CMIP6 group (Meehl *et al.*, 2020). Ensemble means for projected MAT and MAP under the different SSPs and periods varied across all the regions. This highlights the need for region-specific scenario selection (based on the provided Table 8) to capture the uncertainty inherent to future projections by general circulation models across the continent. If computational resources are limited, then users can still select a small number of projections (even two) to quantify uncertainties. One setback to using very small numbers of future projections is that it excludes high sensitivity models such as CanESM1 and UKESM1-0-LL, since they do not represent very likely outcomes. On the other hand, the use of high sensitivity models has been discouraged for the same reason: they may be outliers that do not represent likely future scenarios (for this purpose, representative scenarios can also be selected with outliers excluded (Table 8 second set). Lastly while the KKZ algorithm optimally captures the extent of climate variability for a given number of scenarios (Gudoshava *et al.*, 2024), one note of caution is that strong auto-correlation among a large set of variables may reduce the dimensionality of multivariate space in a way that weighs unique variables (that are not correlated to any other variables) disproportionately high when selecting representative scenarios (Seo *et al.*, 2019). However, in

my analysis I used a balanced set of seasonal and monthly temperature and precipitation variables where this should not be a factor (also supported by relatively moderate and uniform variable loadings in Table 9).

6. Conclusion

In this study, I developed a high-resolution climate database for Africa, addressing the need for detailed and accessible climate data. To achieve this, I compiled a comprehensive monthly weather station database from seven global sources. I interpolated this weather station data to form climate surfaces using three-dimensional (latitude, longitude and elevation) thin-plate spline methodology for monthly precipitation and temperature variables for the 1961-1990 climate normal period at 2.5 arcminute resolution. I then fine-tuned the interpolation using neural networks and a number of covariates including topographic indices, windward distances to coasts and lakes at different resolutions to improve the representation of local climate patterns. These surfaces were cross-validated using the checkerboard method that reduces dependence of training and validation sets. The surfaces were integrated into ClimateAF software package (freely available at <http://tinyurl.com/ClimateAF>) to further downscale climate surfaces using elevation adjustments to produce temperature estimates of higher accuracy and scale-free resolution (grids can be generated at any spatial scale by a user-provided digital elevation model). The surfaces were validated for Africa's IPCC climate regions to assess region-specific accuracy of the layers. Finally I recommended a selection of CMIP6 models for each of the regions that best represents climate uncertainty while minimizing bias using the KKZ algorithm.

The results demonstrated that the neural network fine-tuning approach is effective in capturing complex physiographical effects of local features and topography, providing more accurate climate surfaces compared to traditional methods. These are further downscaled in the ClimateAF software. In addition to these base layers, the software package provides over 24,000 climate surfaces at different temporal scales for standard and bioclimatic variables from 1901-2023, as well as CMIP6 future climate projections until 2100. The easy access of high resolution climate data through this software should make it a valuable resource for researchers in climate

science, ecology, forestry and other fields as well as decision-makers to advance knowledge and climate adaptation efforts in Africa.

References

- Allaire, J.J. & Chollet, F. (2023) keras: R Interface to 'Keras'. In: *R package version 2.13.0*
- Alsafadi, K., Bi, S., Bashir, B., Sharifi, E., Alsaman, A., Kumar, A. & Shahid, S. (2023) High-Resolution Precipitation Modeling in Complex Terrains Using Hybrid Interpolation Techniques: Incorporating Physiographic and MODIS Cloud Cover Influences. In: *Remote Sensing*
- Alvares, C.A., Stape, J.L., Sentelhas, P.C. & Gonçalves, J.L.D. (2013) Modeling monthly mean air temperature for Brazil. *Theoretical and Applied Climatology*, **113**, 407-427.
- Antonić, O., Križan, J., Marki, A. & Bukovec, D. (2001) Spatio-temporal interpolation of climatic variables over large region of complex terrain using neural networks. *Ecological Modelling*, **138**, 255-263.
- Arguez, A. & Vose, R.S. (2011) The Definition of the Standard WMO Climate Normal: The Key to Deriving Alternative Climate Normals. *Bulletin of the American Meteorological Society*, **92**, 699-704.
- Attorre, F., Alfo, M., De Sanctis, M., Francesconi, F. & Bruno, F. (2007) Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale. *International Journal of Climatology*, **27**, 1825-1843.
- Barsugli, J.J., Guentchev, G., Horton, R.M., Wood, A., Mearns, L.O., Liang, X.-Z., Winkler, J.A., Dixon, K., Hayhoe, K., Rood, R.B., Goddard, L., Ray, A., Buja, L. & Ammann, C. (2013) The Practitioner's Dilemma: How to Assess the Credibility of Downscaled Climate Projections. *Eos, Transactions American Geophysical Union*, **94**, 424-425.
- Biecek, P. (2018) DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*, **19**, 1-5.
- Bliefernicht, J., Salack, S., Waongo, M., Annor, T., Laux, P. & Kunstmann, H. (2022) Towards a historical precipitation database for West Africa: Overview, quality control and harmonization. *International Journal of Climatology*, **42**, 4001-4023.
- Boer, E.P.J., de Beurs, K.M. & Hartkamp, A.D. (2001) Kriging and thin plate splines for mapping climate variables. *International Journal of Applied Earth Observation and Geoinformation*, **3**, 146-154.
- Booth, T. (1985) A new method for assisting species selection. *The Commonwealth Forestry Review*, 241-250.
- Boucher, O., Servonnat, J., Albright, A.L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, L.E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luysaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A.K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N. & Vuichard, N. (2020) Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS002010.

- Bradley St. Clair, J., Kilkenny, F.F., Johnson, R.C., Shaw, N.L. & Weaver, G. (2013) Genetic variation in adaptive traits and seed transfer zones for *Pseudoroegneria spicata* (bluebunch wheatgrass) in the northwestern United States. *Evolutionary Applications*, **6**, 933-948.
- Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32.
- Cannon, A.J. (2015) Selecting GCM Scenarios that Span the Range of Changes in a Multimodel Ensemble: Application to CMIP5 Climate Extremes Indices. *Journal of Climate*, **28**, 1260-1267.
- Cansler, C.A., Kane, V.R., Hessburg, P.F., Kane, J.T., Jeronimo, S.M., Lutz, J.A., Povak, N.A., Churchill, D.J. & Larson, A.J. (2022) Previous wildfires and management treatments moderate subsequent fire severity. *Forest Ecology and Management*, **504**, 119764.
- Castellanos-Acuna, D. & Hamann, A. (2020) A cross-checked global monthly weather station database for precipitation covering the period 1901–2010. *Geoscience Data Journal*, **7**, 27-37.
- Chiles, J.-P. & Delfiner, P. (2012) Kriging. *Geostatistics: modeling spatial uncertainty*, pp. 147-237. John Wiley & Sons.
- Chollet, F. (2022) *Deep learning with R*, Second edition. edn. Manning.
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W.J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A.J. & Wehner, M. (2013) Long-term Climate Change: Projections, Commitments and Irreversibility. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley), pp. 1029–1136. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M. & Jones, P.D. (2018) An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, **123**, 9391-9409.
- Cortini, F., Comeau, P.G. & Bokalo, M. (2012) Trembling aspen competition and climate effects on white spruce growth in boreal mixtures of Western Canada. *Forest Ecology and Management*, **277**, 67-73.
- Cressie, N.A. (1993) Statistics for spatial data. John Willy and Sons. *Inc.*, New York, **800**
- Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J. & Pasteris, P.P. (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, **28**, 2031-2064.
- De La Torre, A.R., Wang, T., Jaquish, B. & Aitken, S.N. (2014) Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: implications for forest management under climate change. *New Phytologist*, **201**, 687-699.
- Deblauwe, V., Droissart, V., Bose, R., Sonké, B., Blach-Overgaard, A., Svenning, J.C., Wieringa, J.J., Ramesh, B.R., Stévant, T. & Couvreur, T.L.P. (2016) Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Global Ecology and Biogeography*, **25**, 443-454.
- DeGaetano, A.T. & Belcher, B.N. (2007) Spatial interpolation of daily maximum and minimum air temperature based on meteorological model analyses and independent observations. *Journal of Applied Meteorology and Climatology*, **46**, 1981-1992.

- 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., Nieradzick, 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. & Zhang, Q. (2022) The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. *Geosci. Model Dev.*, **15**, 2973-3020.
- Dunne, J.P., Horowitz, L.W., Adcroft, A.J., Ginoux, P., Held, I.M., John, J.G., Krasting, J.P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C.A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K.A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P.P.G., Griffies, S.M., Guo, H., Hallberg, R.W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly, P.C.D., Nikonov, S., Paynter, D.J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B.G., Robinson, T., Schwarzkopf, D.M., Sentman, L.T., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A.T., Wyman, B., Zeng, Y. & Zhao, M. (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.
- Elhakeem, A., Hughes, R.A., Tilling, K., Cousminer, D.L., Jackowski, S.A., Cole, T.J., Kwong, A.S.F., Li, Z., Grant, S.F.A., Baxter-Jones, A.D.G., Zemel, B.S. & Lawlor, D.A. (2022) Using linear and natural cubic splines, SITAR, and latent trajectory models to characterise nonlinear longitudinal growth trajectories in cohort studies. *BMC Medical Research Methodology*, **22**, 68.
- Endris, H.S., Omondi, P., Jain, S., Lennard, C., Hewitson, B., Chang'a, L., Awange, J.L., Dosio, A., Ketiemi, P., Nikulin, G., Panitz, H.J., Buchner, M., Stordal, F. & Tazalika, L. (2013) Assessment of the Performance of CORDEX Regional Climate Models in Simulating East African Rainfall. *Journal of Climate*, **26**, 8453-8475.
- Engelbrecht, F., Adegoke, J., Bopape, M.-J., Naidoo, M., Garland, R., Thatcher, M., McGregor, J., Katzfey, J., Werner, M., Ichoku, C. & Gatebe, C. (2015) Projections of rapidly rising surface temperatures over Africa under low mitigation. *Environmental Research Letters*, **10**, 085004.
- ESRI (2008) *ArcGIS Desktop: Release 9.3*
- ESRI (2019) EM SHR Station Locations and Metadata 1738 to Present. In: ArcGIS Online
- 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. *Geosci. Model Dev.*, **9**, 1937-1958.
- FAO (2001) FAOCLIM 2: world-wide agroclimatic data. In: *Environment and Natural Resources, Working paper No. 5*
- Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302-4315.
- 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.
- Funk, C.C., Peterson, P.J., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J., Michaelsen, J.C. & Verdin, A.P. (2014) A quasi-global precipitation time series for drought monitoring: U.S. Geological Survey Data Series 832. 4.
- Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Gi-Kong, K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M. & Zhao, B. (2017) The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, **30**, 5419-5454.
- Georgeson, L., Maslin, M. & Poessinouw, M. (2017) Global disparity in the supply of commercial weather and climate information services. *Science Advances*, **3**, e1602632.
- Gesch, D.B., Verdin, K.L. & Greenlee, S.K. (1999) New land surface digital elevation model covers the Earth. *Eos, Transactions American Geophysical Union*, **80**, 69-70.
- Gleckler, P.J., Taylor, K.E. & Doutriaux, C. (2008) Performance metrics for climate models. *Journal of Geophysical Research: Atmospheres*, **113**
- Gudoshava, M., Otieno, G., Koech, E., Misiani, H., Ongoma, J.G., Heinrich-Mertsching, C., Wachana, C., Endris, H.S., Mwanthi, A., Kilavi, M., Mwangi, E., Colman, A., Parker, D., Mutemi, J.N., Machio, P., Omay, P.O., Ombai, P., Anande, D., Kondowe, A., Mugume, I., Ayabagabo, P., Houssein, H.Y., Waiss, M.S., Abeshu, B., Kayoya, E., Sharawe, M.N., Bahaga, T., Todd, M., Segele, Z., Atheru, Z. & Artan, G. (2024) Advances, gaps and way forward in provision of climate services over the Greater Horn of Africa. *Frontiers in Climate*, **6**
- Guttman, N.B. (1989) Statistical Descriptors of Climate. *Bulletin of the American Meteorological Society*, **70**, 602-607.
- Harris, I., Jones, P.D., Osborn, T.J. & Lister, D.H. (2014) Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology*, **34**, 623-642.
- Hartkamp, A.D., De Beurs, K., Stein, A. & White, J.W. (1999) Interpolation Techniques for Climate Variables. In: *CIMMYT NRG-GIS Series*, pp. 1-34, Mexico City.
- Hartmann, D.L., Klein Tank, A.M.G., Rusticucci, M., Alexander, L.V., Brönnimann, S., Charabi, Y., Dentener, F.J., Dlugokencky, E.J., Easterling, D.R., Kaplan, A., Soden, B.J., Thorne, P.W., Wild, M. & Zhai, P.M. (2013) Observations: Atmosphere and Surface. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley), pp. 159–254. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Hawkins, D.M. & Cressie, N. (1984) Robust kriging—A proposal. *Journal of the International Association for Mathematical Geology*, **16**, 3-18.
- Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M. & Gräler, B. (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *Peerj*, **6**

- Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., de Jesus, J.M., Tamene, L. & Tondoh, J.E. (2015) Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *Plos One*, **10**
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. & Thépaut, J.-N. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999-2049.
- Hewitt, C.D., Allis, E., Mason, S.J., Muth, M., Pulwarty, R., Shumake-Guillemot, J., Bucher, A., Brunet, M., Fischer, A.M., Hama, A.M., Kolli, R.K., Lucio, F., Ndiaye, O. & Tapia, B. (2020) Making Society Climate Resilient: International Progress under the Global Framework for Climate Services. *Bulletin of the American Meteorological Society*, **101**, E237-E252.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.
- Hoaglin, D.C., Mosteller, F. & Tukey, J.W. (2000) *Understanding robust and exploratory data analysis*. John Wiley & Sons.
- Hoerling, M., Hurrell, J., Eischeid, J. & Phillips, A. (2006) Detection and Attribution of Twentieth-Century Northern and Southern African Rainfall Change. *Journal of Climate*, **19**, 3989-4008.
- Hoerling, M., Hurrell, J., Kumar, A., Terray, L., Eischeid, J., Pegion, P., Zhang, T., Quan, X. & Xu, T. (2011) On North American Decadal Climate for 2011–20. *Journal of Climate*, **24**, 4519-4528.
- Houle, D., Bouffard, A., Duchesne, L., Logan, T. & Harvey, R. (2012) Projections of Future Soil Temperature and Water Content for Three Southern Quebec Forested Sites. *Journal of Climate*, **25**, 7690-7701.
- Hu, X., Hou, Y., Li, D., Hua, T., Marchi, M., Paola Forero Urrego, J., Huang, B., Zhao, W. & Cherubini, F. (2023) Changes in multiple ecosystem services and their influencing factors in Nordic countries. *Ecological Indicators*, **146**, 109847.
- Hutchinson, M.F. (1995a) Interpolating mean rainfall using thin plate smoothing splines. *International journal of geographical information systems*, **9**, 385-403.
- Hutchinson, M.F. & Gessler, P.E. (1994) Splines — more than just a smooth interpolator. *Geoderma*, **62**, 45-67.
- Hutchinson, M.F. & Xu, T. (2013) ANUSPLIN Version 4.4 User Guide. In. Australian National University - Fenner School of Environment & Society
- Illés, G. & Mórnicz, N. (2022) Climate envelope analyses suggests significant rearrangements in the distribution ranges of Central European tree species. *Annals of Forest Science*, **79**, 35.
- IPCC (2021) *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom; New York, NY, USA.

- Iturbide, M., Gutiérrez, J.M., Alves, L.M., Bedia, J., Cerezo-Mota, R., Gimenez, E., Cofiño, A.S., Di Luca, A., Faria, S.H., Gorodetskaya, I.V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H.T., Jones, R.G., Krakovska, S., Manzanar, R., Martínez-Castro, D., Narisma, G.T., Nurhati, I.S., Pinto, I., Seneviratne, S.I., van den Hurk, B. & Vera, C.S. (2020) An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets. *Earth Syst. Sci. Data*, **12**, 2959-2970.
- Jarnevich, C.S., Young, N.E., Talbert, M. & Talbert, C. (2018) Forecasting an invasive species' distribution with global distribution data, local data, and physiological information. *Ecosphere*, **9**, e02279.
- Jiang, Y. (2009) Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models. *Energy*, **34**, 1276-1283.
- Jones, P.D., New, M., Parker, D.E., Martin, S. & Rigor, I.G. (1999) Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, **37**, 173-199.
- Karger, D.N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P. & Kessler, M. (2017) Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, **4**, 170122.
- Karmalkar, A.V., Thibeault, J.M., Bryan, A.M. & Seth, A. (2019) Identifying credible and diverse GCMs for regional climate change studies—case study: Northeastern United States. *Climatic Change*, **154**, 367-386.
- Katsavounidis, I., Kuo, C.C.J. & Zhang, Z. (1994) A new initialization technique for generalized Lloyd iteration. *IEEE Signal Processing Letters*, **1**, 144-146.
- Kelley, M., Schmidt, G.A., Nazarenko, L.S., Bauer, S.E., Ruedy, R., Russell, G.L., Ackerman, A.S., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Cesana, G., Cheng, Y., Clune, T.L., Cook, B.I., Cruz, C.A., Del Genio, A.D., Elsaesser, G.S., Faluvegi, G., Kiang, N.Y., Kim, D., Lacis, A.A., Leboissetier, A., LeGrande, A.N., Lo, K.K., Marshall, J., Matthews, E.E., McDermid, S., Mezuman, K., Miller, R.L., Murray, L.T., Oinas, V., Orbe, C., García-Pando, C.P., Perlwitz, J.P., Puma, M.J., Rind, D., Romanou, A., Shindell, D.T., Sun, S., Tausnev, N., Tsigaridis, K., Tselioudis, G., Weng, E., Wu, J. & Yao, M.-S. (2020) GISS-E2.1: Configurations and Climatology. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS002025.
- Kirtman, B., Power, S.B., Adedoyin, J.A., Boer, G.J., Bojariu, R., Camilloni, I., Doblas-Reyes, F.J., Fiore, A.M., Kimoto, M., Meehl, G.A., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G.J., Vecchi, G. & Wang, H.J. (2013) Near-term Climate Change: Projections and Predictability. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley), pp. 953–1028. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, G.A. (2010) Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate*, **23**, 2739-2758.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. (2006) World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, **15**, {259-263}.
- Lawrimore, J.H., Menne, M.J., Gleason, B.E., Williams, C.N., Wuertz, D.B., Vose, R.S. & Rennie, J. (2011) An overview of the Global Historical Climatology Network monthly

- mean temperature data set, version 3. *Journal of Geophysical Research-Atmospheres*, **116**
- Lee, J.-K. & Kim, Y.-O. (2017) Selection of representative GCM scenarios preserving uncertainties. *Journal of Water and Climate Change*, **8**, 641-651.
- Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J.P., Engelbrecht, F., Fischer, E., Fyfe, J.C., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S. & Zhou, T. (2021) Future Global Climate: Scenario-based Projections and Near-term Information. *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by C. Intergovernmental Panel on Climate), pp. 553-672. Cambridge University Press, Cambridge.
- Lee, T.C.K., Zwiers, F.W., Zhang, X. & Tsao, M. (2006) Evidence of Decadal Climate Prediction Skill Resulting from Changes in Anthropogenic Forcing. *Journal of Climate*, **19**, 5305-5318.
- Li, J. & Heap, A.D. (2008) *A review of spatial interpolation methods for environmental scientists*. Geoscience Australia, Australian Government Canberra, Australia.
- Li, J. & Heap, A.D. (2014) Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, **53**, 173-189.
- Liang, Y.X., Gillett, N.P. & Monahan, A.H. (2020) Climate Model Projections of 21st Century Global Warming Constrained Using the Observed Warming Trend. *Geophysical Research Letters*, **47**
- Lovell, J.T., MacQueen, A.H., Mamidi, S., Bonnette, J., Jenkins, J., Napier, J.D., Sreedasyam, A., Healey, A., Session, A., Shu, S., Barry, K., Bonos, S., Boston, L., Daum, C., Deshpande, S., Ewing, A., Grabowski, P.P., Haque, T., Harrison, M., Jiang, J., Kudrna, D., Lipzen, A., Pendergast, T.H., Plott, C., Qi, P., Saski, C.A., Shakirov, E.V., Sims, D., Sharma, M., Sharma, R., Stewart, A., Singan, V.R., Tang, Y., Thibivillier, S., Webber, J., Weng, X., Williams, M., Wu, G.A., Yoshinaga, Y., Zane, M., Zhang, L., Zhang, J., Behrman, K.D., Boe, A.R., Fay, P.A., Fritschi, F.B., Jastrow, J.D., Lloyd-Reilley, J., Martínez-Reyna, J.M., Matamala, R., Mitchell, R.B., Rouquette, F.M., Ronald, P., Saha, M., Tobias, C.M., Udvardi, M., Wing, R.A., Wu, Y., Bartley, L.E., Casler, M., Devos, K.M., Lowry, D.B., Rokhsar, D.S., Grimwood, J., Juenger, T.E. & Schmutz, J. (2021) Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature*, **590**, 438-444.
- Mahony, C.R., Wang, T.L., Hamann, A. & Cannon, A.J. (2022) A global climate model ensemble for downscaled monthly climate normals over North America. *International Journal of Climatology*, **42**, 5871-5891.
- Maidment, R.I., Grimes, D., Allan, R.P., Tarnavsky, E., Stringer, M., Hewison, T., Roebeling, R. & Black, E. (2014) The 30 year TAMSAT African Rainfall Climatology And Time series (TARCAT) data set. *Journal of Geophysical Research: Atmospheres*, **119**, 10,619-10,644.
- Maier, H.R. & Dandy, G.C. (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, **15**, 101-124.
- Mandler, H. & Weigand, B. (2023) Feature importance in neural networks as a means of interpretation for data-driven turbulence models. *Computers & Fluids*, **265**, 105993.

- Marchi, M., Sinjur, I., Bozzano, M. & Westergren, M. (2019) Evaluating WorldClim Version 1 (1961–1990) as the Baseline for Sustainable Use of Forest and Environmental Resources in a Changing Climate. In: *Sustainability*
- Marchi, M., Castellanos-Acuña, D., Hamann, A., Wang, T., Ray, D. & Menzel, A. (2020) ClimateEU, scale-free climate normals, historical time series, and future projections for Europe. *Scientific Data*, **7**, 428.
- Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K. & Lonnoy, E. (2021) 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. *IPCC* (ed. by J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou), pp. 2087-2138. Cambridge University Press.
- Masson, D. & Knutti, R. (2011) Climate model genealogy. *Geophysical Research Letters*, **38**
- Matheron, G. (1963) Principles of geostatistics. *Economic Geology*, **58**, 1246-1266.
- Matheron, G. (1971) *The Theory of Regionalized Variables and Its Applications*. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Paris, France.
- Mathur, M., Mathur, P. & Purohit, H. (2023) Ecological niche modelling of a critically endangered species *Commiphora wightii* (Arn.) Bhandari using bioclimatic and non-bioclimatic variables. *Ecological Processes*, **12**, 8.
- Mbogga, M.S., Hansen, S., Wang, T. & Hamann, A. (2010) *A comprehensive set of interpolated climate data for Alberta*. . Government of Alberta, Alberta Sustainable Resource Development.
- McCulloch, W.S. & Pitts, W. (1990) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, **52**, 99-115.
- McSweeney, C.F., Jones, R.G., Lee, R.W. & Rowell, D.P. (2015) Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dynamics*, **44**, 3237-3260.
- Meehl, G.A., Senior, C.A., Eyring, V., Flato, G., Lamarque, J.F., Stouffer, R.J., Taylor, K.E. & Schlund, M. (2020) Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, **6**
- Meinshausen, M., Nicholls, Z.R.J., Lewis, J., Gidden, M.J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J.G., Daniel, J.S., John, A., Krummel, P.B., Luderer, G., Meinshausen, N., Montzka, S.A., Rayner, P.J., Reimann, S., Smith, S.J., van den Berg, M., Velders, G.J.M., Vollmer, M.K. & Wang, R.H.J. (2020) The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geosci. Model Dev.*, **13**, 3571-3605.
- Mejia, J.M. & Rousselle, J. (1976) Disaggregation models in hydrology revisited. *Water Resources Research*, **12**, 185-186.
- Mekonnen, Z.A., Riley, W.J., Randerson, J.T., Grant, R.F. & Rogers, B.M. (2019) Expansion of high-latitude deciduous forests driven by interactions between climate warming and fire. *Nature plants*, **5**, 952-958.
- Mendlik, T. & Gobiet, A. (2016) Selecting climate simulations for impact studies based on multivariate patterns of climate change. *Climatic Change*, **135**, 381-393.
- Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E. & Houston, T.G. (2012) An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*, **29**, 897-910.

- Minaei, A., Todeschini, S., Sitzenfrei, R. & Creaco, E. (2022) Ensemble Evaluation and Member Selection of Regional Climate Models for Impact Models Assessment. In: *Water*
- Mote, P.W., Allen, M.R., Jones, R.G., Li, S.H., Mera, R., Rupp, D.E., Salahuddin, A. & Vickers, D. (2016) Superensemble Regional Climate Modeling for the Western United States. *Bulletin of the American Meteorological Society*, **97**, 203-216.
- Müller, W.A., Jungclaus, J.H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornbluh, L., Li, H., Modali, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., Tian, F. & Marotzke, J. (2018) A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, **10**, 1383-1413.
- Nalder, I.A. & Wein, R.W. (1998) Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest. *Agricultural and Forest Meteorology*, **92**, 211-225.
- New, M., Hulme, M. & Jones, P. (1999) Representing Twentieth-Century space-time climate variability. Part I: Development of a 1961-90 mean monthly terrestrial climatology. *Journal of Climate*, **12**, 829-856.
- Ngarega, B.K., Chaibva, P., Masocha, V.F., Saina, J.K., Khine, P.K. & Schneider, H. (2023) Application of MaxEnt modeling to evaluate the climate change effects on the geographic distribution of *Lippia javanica* (Burm.f.) Spreng in Africa. *Environmental Monitoring and Assessment*, **196**, 62.
- Nicholson, S.E. (2013) The West African Sahel: A Review of Recent Studies on the Rainfall Regime and Its Interannual Variability. *ISRN Meteorology*, **2013**, 453521.
- NOAA (2018) Federal climate complex data documentation for integrated surface data (ISD). In. NOAA - National Centers for Environmental Information, Asheville, NC, USA.
- Novella, N.S. & Thiaw, W.M. (2013) African Rainfall Climatology Version 2 for Famine Early Warning Systems. *Journal of Applied Meteorology and Climatology*, **52**, 588-606.
- Nychka, D., Furrer, R., Paige, J. & Sain, S. (2021) fields: Tools for spatial data. In: *R package version 15.2*. University Corporation for Atmospheric Research, Boulder, CO, USA.
- Ogougbé, R.E., Agbo, R.I., Ahamidé, B., Djego-Djossou, S. & Djego, G.J. (2022) Potential geographic distribution and modelling of the ecological niche of *Harrisonia abyssinica*, a priority medicinal tree species in Benin (West Africa). *Modeling Earth Systems and Environment*, **8**, 2471-2483.
- Omukuti, J., Wanzala, M.A., Ngaina, J. & Ganola, P. (2023) Develop medium- to long-term climate information services to enhance comprehensive climate risk management in Africa. *Climate Resilience and Sustainability*, **2**
- Overland, J.E., Wang, M.Y., Walsh, J.E. & Stroeve, J.C. (2014) Future Arctic climate changes: Adaptation and mitigation time scales. *Earths Future*, **2**, 68-74.
- Pang, B., Yue, J., Zhao, G. & Xu, Z. (2017) Statistical Downscaling of Temperature with the Random Forest Model. *Advances in Meteorology*, **2017**, 7265178.
- Parisien, M.-A., Barber, Q.E., Bourbonnais, M.L., Daniels, L.D., Flannigan, M.D., Gray, R.W., Hoffman, K.M., Jain, P., Stephens, S.L. & Taylor, S.W. (2023) Abrupt, climate-induced increase in wildfires in British Columbia since the mid-2000s. *Communications Earth & Environment*, **4**, 309.
- Parks, S.A., Holsinger, L.M., Panunto, M.H., Jolly, W.M., Dobrowski, S.Z. & Dillon, G.K. (2018) High-severity fire: evaluating its key drivers and mapping its probability across western US forests. *Environmental research letters*, **13**, 044037.

- PCMDI, P.f.C.M.D.I. (2022) *The Program for Climate Model Diagnosis and Intercomparison: furthering the understanding of climate change through model and observational analysis and community leadership*. Available at: <https://pcmdi.llnl.gov/about.html> (accessed April, 20 2024).
- PCMDI, P.f.C.M.D.I. (2024) *ESGF CMIP6 Data Holdings*. Available at: https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf_data_holdings/index.html (accessed April, 20 2024).
- Räsänen, J. & Ruokolainen, L. (2006) Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus A: Dynamic Meteorology and Oceanography*, **58**, 461-472.
- Rampal, N., Gibson, P.B., Sood, A., Stuart, S., Fauchereau, N.C., Brandolino, C., Noll, B. & Meyers, T. (2022) High-resolution downscaling with interpretable deep learning: Rainfall extremes over New Zealand. *Weather and Climate Extremes*, **38**, 100525.
- Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fife, J., Fife, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A. & Taylor, K.E. (2007) Climate Models and Their Evaluation. *The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Rigol, J.P., Jarvis, C.H. & Stuart, N. (2001) Artificial neural networks as a tool for spatial interpolation. *International Journal of Geographical Information Science*, **15**, 323-343.
- Rubel, F. & Kottek, M. (2010) Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift*, **19**, {135-141}.
- Sáenz-Romero, C., O'Neill, G., Aitken, S.N. & Lindig-Cisneros, R. (2020) Assisted migration field tests in Canada and Mexico: Lessons, limitations, and challenges. *Forests*, **12**, 9.
- Schirpke, U. & Ebner, M. (2022) Exposure to global change pressures and potential impacts on ecosystem services of mountain lakes in the European Alps. *Journal of Environmental Management*, **318**, 115606.
- Schlenker, W., Hanemann, W.M. & Fisher, A.C. (2007) Water Availability, Degree Days, and the Potential Impact of Climate Change on Irrigated Agriculture in California. *Climatic Change*, **81**, 19-38.
- Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voltaire, A., Colin, J., Decharme, B., Delire, C., Berthet, S., Chevallier, M., Sénési, S., Franchisteguy, L., Vial, J., Mallet, M., Joetzjer, E., Geoffroy, O., Guérémy, J.-F., Moine, M.-P., Msadek, R., Ribes, A., Rocher, M., Roehrig, R., Salas-y-Méla, D., Sanchez, E., Terray, L., Valcke, S., Waldman, R., Aumont, O., Bopp, L., Deshayes, J., Éthé, C. & Madec, G. (2019) Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate. *Journal of Advances in Modeling Earth Systems*, **11**, 4182-4227.
- Sekulić, A., Kilibarda, M., Heuvelink, G.B.M., Nikolić, M. & Bajat, B. (2020) Random Forest Spatial Interpolation. In: *Remote Sensing*
- Sellar, A.A., Jones, C.G., Mulcahy, J.P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F.M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S.T., Kelley, D.I., Ellis, R., Johnson, C.E., Walton, J., Abraham, N.L., Andrews, M.B., Andrews, T., Archibald, A.T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi,

- M., Edwards, J., Folberth, G.A., Gedney, N., Griffiths, P.T., Harper, A.B., Hendry, M.A., Hewitt, A.J., Johnson, B., Jones, A., Jones, C.D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R.J., Predoi, V., Robertson, E., Siahaan, A., Smith, R.S., Swaminathan, R., Woodhouse, M.T., Zeng, G. & Zerroukat, M. (2019) UKESM1: Description and Evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*, **11**, 4513-4558.
- Seo, S.B., Kim, Y.O., Kim, Y. & Eum, H.I. (2019) Selecting climate change scenarios for regional hydrologic impact studies based on climate extremes indices. *Climate Dynamics*, **52**, 1595-1611.
- Shepard, D. (1968) A two-dimensional interpolation function for irregularly-spaced data. In: *Proceedings of the 1968 23rd ACM national conference*, pp. 517–524. Association for Computing Machinery
- Shrestha, A.B., Wake, C.P., Mayewski, P.A. & Dibb, J.E. (1999) Maximum Temperature Trends in the Himalaya and Its Vicinity: An Analysis Based on Temperature Records from Nepal for the Period 1971–94. *Journal of Climate*, **12**, 2775-2786.
- Smith, K. (1975) *Principles of Applied Climatology*. Wiley.
- Snell, S.E., Gopal, S. & Kaufmann, R.K. (2000) Spatial Interpolation of Surface Air Temperatures Using Artificial Neural Networks: Evaluating Their Use for Downscaling GCMs. *Journal of Climate*, **13**, 886-895.
- Sung, J.H., Kwon, M., Jeon, J.-J. & Seo, S.B. (2019) A Projection of Extreme Precipitation Based on a Selection of CMIP5 GCMs over North Korea. In: *Sustainability*
- Suzuki, R., Terada, Y. & Shimodaira, H. (2019) *pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling*.
- 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., Sigmund, M., Solheim, L., von Salzen, K., Yang, D. & Winter, B. (2019) The Canadian Earth System Model version 5 (CanESM5.0.3). *Geosci. Model Dev.*, **12**, 4823-4873.
- Tan, J., Xie, X., Zuo, J., Xing, X., Liu, B., Xia, Q. & Zhang, Y. (2021) Coupling random forest and inverse distance weighting to generate climate surfaces of precipitation and temperature with Multiple-Covariates. *Journal of Hydrology*, **598**, 126270.
- Tank, A.M.G.K., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., Van Engelen, A.F.V., Forland, E., Miletus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., López, J.A., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L.V. & Petrovic, P. (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology*, **22**, 1441-1453.
- Tarboton, D.G., Sharma, A. & Lall, U. (1998) Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water Resources Research*, **34**, 107-119.
- 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. & Kimoto, M. (2019) Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.*, **12**, 2727-2765.
- Taylor, K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, **106**, 7183-7192.
- Tett, S.F.B., Stott, P.A., Allen, M.R., Ingram, W.J. & Mitchell, J.F.B. (1999) Causes of twentieth-century temperature change near the Earth's surface. *Nature*, **399**, 569-572.
- Thyer, M. & Kuczera, G. (2000) Modeling long-term persistence in hydroclimatic time series using a hidden state Markov Model. *Water Resources Research*, **36**, 3301-3310.
- USGS, U.S.G.S. (1996) GTOPO30, A global digital elevation model (DEM). In:
- van Niekerk, A. & Joubert, S.J. (2011) Input variable selection for interpolating high-resolution climate surfaces for the Western Cape. *Water SA*, **37**, 271-279.
- Verdin, A., Funk, C., Peterson, P., Landsfeld, M., Tuholske, C. & Grace, K. (2020) Development and validation of the CHIRTS-daily quasi-global high-resolution daily temperature data set. *Scientific Data*, **7**, 303.
- Volodin, E.M., Mortikov, E.V., Kostykin, S.V., Galin, V.Y., Lykossov, V.N., Gritsun, A.S., Diansky, N.A., Gusev, A.V. & Iakovlev, N.G. (2017) Simulation of the present-day climate with the climate model INMCM5. *Climate Dynamics*, **49**, 3715-3734.
- Vose, R.S., McNeill, S., Thomas, K. & Shepherd, E. (2011) Enhanced Master Station History Report of March 2019. In. NOAA National Climatic Data Center
- Wahba, G. (1979) How to Smooth Curves and Surfaces with Splines and Cross-Validation. In. Wisconsin University - Madison Department of Statistics
- Wang, T., Hamann, A., Spittlehouse, D.L. & Aitken, S.N. (2006) Development of scale-free climate data for western Canada for use in resource management. *Int. J. Climatol.*, **26**, 383-397.
- Wang, T., Hamann, A., Spittlehouse, D.L. & Murdock, T.Q. (2012) ClimateWNA—High-Resolution Spatial Climate Data for Western North America. *Journal of Applied Meteorology and Climatology*, **51**, 16-29.
- Wang, T.L., Hamann, A., Spittlehouse, D. & Carroll, C. (2016) Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America. *Plos One*, **11**
- Weigel, A.P., Knutti, R., Liniger, M.A. & Appenzeller, C. (2010) Risks of Model Weighting in Multimodel Climate Projections. *Journal of Climate*, **23**, 4175-4191.
- Wesselkamp, M., Roberts, D.R. & Dormann, C.F. (2024) Identifying potential provenances for climate-change adaptation using spatially variable coefficient models. *BMC Ecology and Evolution*, **24**, 70.
- Whittleston, D., Nicholson, S.E., Schlosser, A. & Entekhabi, D. (2017) Climate Models Lack Jet-Rainfall Coupling over West Africa. *Journal of Climate*, **30**, 4625-4632.
- Wilcke, R.A.I. & Barring, L. (2016) Selecting regional climate scenarios for impact modelling studies. *Environmental Modelling & Software*, **78**, 191-201.
- Wilks, D.S. (2019) Chapter 10 - Time Series. *Statistical Methods in the Atmospheric Sciences (Fourth Edition)* (ed. by D.S. Wilks), pp. 485-550. Elsevier.
- Willmott, C.J. & Matsuura, K. (1995) Smart Interpolation of Annually Averaged Air Temperature in the United States. *Journal of Applied Meteorology (1988-2005)*, **34**, 2577-2586.

- WMO (1996) Climatological Normals (CLINO) for the period 1961-1990. In, p. 788. World Meteorological Organization, Geneva, Switzerland.
- WMO, W.M.O. (2022) *Essential Climate Variables*. Available at: <https://public.wmo.int/en/programmes/global-climate-observing-system/essential-climate-variables> (accessed April 01 2023)
- Wolfensberger, D., Gabella, M., Boscacci, M., Germann, U. & Berne, A. (2021) RainForest: a random forest algorithm for quantitative precipitation estimation over Switzerland. *Atmospheric Measurement Techniques*, **14**, 3169-3193.
- Workneh, H.T., Chen, X., Ma, Y., Bayable, E. & Dash, A. (2024) Comparison of IDW, Kriging and orographic based linear interpolations of rainfall in six rainfall regimes of Ethiopia. *Journal of Hydrology: Regional Studies*, **52**, 101696.
- Wu, T. & Li, Y. (2013) Spatial interpolation of temperature in the United States using residual kriging. *Applied Geography*, **44**, 112-120.
- Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., Jie, W., Zhang, J., Liu, Y., Zhang, L., Zhang, F., Zhang, Y., Wu, F., Li, J., Chu, M., Wang, Z., Shi, X., Liu, X., Wei, M., Huang, A., Zhang, Y. & Liu, X. (2019) The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6. *Geosci. Model Dev.*, **12**, 1573-1600.
- Xie, P. & Arkin, P.A. (1997) Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs. *Bulletin of the American Meteorological Society*, **78**, 2539-2558.
- Yatagai, A., Arakawa, O., Kamiguchi, K., Kawamoto, H., Nodzu, M.I. & Hamada, A. (2010) A 44-year daily gridded precipitation dataset for Asia based on a dense network of rain gauges, SOLA. In:
- Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N. & Kitoh, A. (2012) APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges. *Bulletin of the American Meteorological Society*, **93**, 1401-1415.
- Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y. & Ishii, M. (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. Ser. II*, **advpub**
- Zeng, Z., Wang, Z., Gui, K., Yan, X., Gao, M., Luo, M., Geng, H., Liao, T., Li, X., An, J., Liu, H., He, C., Ning, G. & Yang, Y. (2020) Daily Global Solar Radiation in China Estimated From High-Density Meteorological Observations: A Random Forest Model Framework. *Earth and Space Science*, **7**, e2019EA001058.
- Zhan, J., Wu, S., Qi, J., Zeng, J., Qin, M., Wang, Y. & Du, Z. (2023) A generalized spatial autoregressive neural network method for three-dimensional spatial interpolation. *Geosci. Model Dev.*, **16**, 2777-2794.
- Ziehn, T., Chamberlain, M.A., Law, R.M., Lenton, A., Bodman, R.W., Dix, M., Stevens, L., Wang, Y.-P. & Srbinovsky, J. (2020) The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, **70**, 193-214.

- Allaire, J.J. & Chollet, F. (2023) keras: R Interface to 'Keras'. In: *R package version 2.13.0*
- Alsafadi, K., Bi, S., Bashir, B., Sharifi, E., Alsalman, A., Kumar, A. & Shahid, S. (2023) High-Resolution Precipitation Modeling in Complex Terrains Using Hybrid Interpolation Techniques: Incorporating Physiographic and MODIS Cloud Cover Influences. In: *Remote Sensing*
- Alvares, C.A., Stape, J.L., Sentelhas, P.C. & Gonçalves, J.L.D. (2013) Modeling monthly mean air temperature for Brazil. *Theoretical and Applied Climatology*, **113**, 407-427.
- Antonić, O., Križan, J., Marki, A. & Bukovec, D. (2001) Spatio-temporal interpolation of climatic variables over large region of complex terrain using neural networks. *Ecological Modelling*, **138**, 255-263.
- Arguez, A. & Vose, R.S. (2011) The Definition of the Standard WMO Climate Normal: The Key to Deriving Alternative Climate Normals. *Bulletin of the American Meteorological Society*, **92**, 699-704.
- Attorre, F., Alfo, M., De Sanctis, M., Francesconi, F. & Bruno, F. (2007) Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale. *International Journal of Climatology*, **27**, 1825-1843.
- Barsugli, J.J., Guentchev, G., Horton, R.M., Wood, A., Mearns, L.O., Liang, X.-Z., Winkler, J.A., Dixon, K., Hayhoe, K., Rood, R.B., Goddard, L., Ray, A., Buja, L. & Ammann, C. (2013) The Practitioner's Dilemma: How to Assess the Credibility of Downscaled Climate Projections. *Eos, Transactions American Geophysical Union*, **94**, 424-425.
- Becker, R.A. (1988) *The New S Language*, 1 edn. Chapman and Hall/CRC, New York.
- Biecek, P. (2018) DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*, **19**, 1-5.
- Bliefernicht, J., Salack, S., Waongo, M., Annor, T., Laux, P. & Kunstmann, H. (2022) Towards a historical precipitation database for West Africa: Overview, quality control and harmonization. *International Journal of Climatology*, **42**, 4001-4023.
- Boer, E.P.J., de Beurs, K.M. & Hartkamp, A.D. (2001) Kriging and thin plate splines for mapping climate variables. *International Journal of Applied Earth Observation and Geoinformation*, **3**, 146-154.
- Booth, T. (1985) A new method for assisting species selection. *The Commonwealth Forestry Review*, 241-250.
- Boucher, O., Servonnat, J., Albright, A.L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, L.E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luysaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A.K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N. & Vuichard, N. (2020) Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS002010.

- Bradley St. Clair, J., Kilkenny, F.F., Johnson, R.C., Shaw, N.L. & Weaver, G. (2013) Genetic variation in adaptive traits and seed transfer zones for *Pseudoroegneria spicata* (bluebunch wheatgrass) in the northwestern United States. *Evolutionary Applications*, **6**, 933-948.
- Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32.
- Cannon, A.J. (2015) Selecting GCM Scenarios that Span the Range of Changes in a Multimodel Ensemble: Application to CMIP5 Climate Extremes Indices. *Journal of Climate*, **28**, 1260-1267.
- Cansler, C.A., Kane, V.R., Hessburg, P.F., Kane, J.T., Jeronimo, S.M., Lutz, J.A., Povak, N.A., Churchill, D.J. & Larson, A.J. (2022) Previous wildfires and management treatments moderate subsequent fire severity. *Forest Ecology and Management*, **504**, 119764.
- Castellanos-Acuna, D. & Hamann, A. (2020) A cross-checked global monthly weather station database for precipitation covering the period 1901-2010. *Geoscience Data Journal*, **7**, 27-37.
- Chiles, J.-P. & Delfiner, P. (2012) Kriging. *Geostatistics: modeling spatial uncertainty*, pp. 147-237. John Wiley & Sons.
- Chollet, F. (2022) *Deep learning with R*, Second edition. edn. Manning.
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W.J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A.J. & Wehner, M. (2013) Long-term Climate Change: Projections, Commitments and Irreversibility. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley), pp. 1029–1136. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M. & Jones, P.D. (2018) An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, **123**, 9391-9409.
- Cortini, F., Comeau, P.G. & Bokalo, M. (2012) Trembling aspen competition and climate effects on white spruce growth in boreal mixtures of Western Canada. *Forest Ecology and Management*, **277**, 67-73.
- Cressie, N.A. (1993) Statistics for spatial data. John Willy and Sons. *Inc.*, New York, **800**
- Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J. & Pasteris, P.P. (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, **28**, 2031-2064.
- De La Torre, A.R., Wang, T., Jaquish, B. & Aitken, S.N. (2014) Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: implications for forest management under climate change. *New Phytologist*, **201**, 687-699.
- Deblauwe, V., Droissart, V., Bose, R., Sonké, B., Blach-Overgaard, A., Svenning, J.C., Wieringa, J.J., Ramesh, B.R., Stévant, T. & Couvreur, T.L.P. (2016) Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Global Ecology and Biogeography*, **25**, 443-454.
- DeGaetano, A.T. & Belcher, B.N. (2007) Spatial interpolation of daily maximum and minimum air temperature based on meteorological model analyses and independent observations. *Journal of Applied Meteorology and Climatology*, **46**, 1981-1992.

- 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., Nieradzick, 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. & Zhang, Q. (2022) The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. *Geosci. Model Dev.*, **15**, 2973-3020.
- Dunne, J.P., Horowitz, L.W., Adcroft, A.J., Ginoux, P., Held, I.M., John, J.G., Krasting, J.P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C.A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K.A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P.P.G., Griffies, S.M., Guo, H., Hallberg, R.W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly, P.C.D., Nikonov, S., Paynter, D.J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B.G., Robinson, T., Schwarzkopf, D.M., Sentman, L.T., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A.T., Wyman, B., Zeng, Y. & Zhao, M. (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.
- Elhakeem, A., Hughes, R.A., Tilling, K., Cousminer, D.L., Jackowski, S.A., Cole, T.J., Kwong, A.S.F., Li, Z., Grant, S.F.A., Baxter-Jones, A.D.G., Zemel, B.S. & Lawlor, D.A. (2022) Using linear and natural cubic splines, SITAR, and latent trajectory models to characterise nonlinear longitudinal growth trajectories in cohort studies. *BMC Medical Research Methodology*, **22**, 68.
- Endris, H.S., Omondi, P., Jain, S., Lennard, C., Hewitson, B., Chang'a, L., Awange, J.L., Dosio, A., Ketiemi, P., Nikulin, G., Panitz, H.J., Buchner, M., Stordal, F. & Tazalika, L. (2013) Assessment of the Performance of CORDEX Regional Climate Models in Simulating East African Rainfall. *Journal of Climate*, **26**, 8453-8475.
- Engelbrecht, F., Adegoke, J., Bopape, M.-J., Naidoo, M., Garland, R., Thatcher, M., McGregor, J., Katzfey, J., Werner, M., Ichoku, C. & Gatebe, C. (2015) Projections of rapidly rising surface temperatures over Africa under low mitigation. *Environmental Research Letters*, **10**, 085004.
- ESRI (2008) *ArcGIS Desktop: Release 9.3*
- ESRI (2019) EM SHR Station Locations and Metadata 1738 to Present. In: ArcGIS Online
- 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. *Geosci. Model Dev.*, **9**, 1937-1958.
- FAO (2001) FAOCLIM 2: world-wide agroclimatic data. In: *Environment and Natural Resources, Working paper No. 5*
- Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302-4315.
- 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.
- Funk, C.C., Peterson, P.J., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J., Michaelsen, J.C. & Verdin, A.P. (2014) A quasi-global precipitation time series for drought monitoring: U.S. Geological Survey Data Series 832. 4.
- Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Gi-Kong, K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M. & Zhao, B. (2017) The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, **30**, 5419-5454.
- Georgeson, L., Maslin, M. & Poessinouw, M. (2017) Global disparity in the supply of commercial weather and climate information services. *Science Advances*, **3**, e1602632.
- Gesch, D.B., Verdin, K.L. & Greenlee, S.K. (1999a) New land surface digital elevation model covers the Earth. *Eos, Transactions American Geophysical Union*, **80**, 69-70.
- Gleckler, P.J., Taylor, K.E. & Doutriaux, C. (2008) Performance metrics for climate models. *Journal of Geophysical Research: Atmospheres*, **113**
- Gudoshava, M., Otieno, G., Koech, E., Misiani, H., Ongoma, J.G., Heinrich-Mertsching, C., Wachana, C., Endris, H.S., Mwanthi, A., Kilavi, M., Mwangi, E., Colman, A., Parker, D., Mutemi, J.N., Machio, P., Omay, P.O., Ombai, P., Anande, D., Kondowe, A., Mugume, I., Ayabagabo, P., Houssein, H.Y., Waiss, M.S., Abeshu, B., Kayoya, E., Sharawe, M.N., Bahaga, T., Todd, M., Segele, Z., Atheru, Z. & Artan, G. (2024) Advances, gaps and way forward in provision of climate services over the Greater Horn of Africa. *Frontiers in Climate*, **6**
- Guttman, N.B. (1989) Statistical Descriptors of Climate. *Bulletin of the American Meteorological Society*, **70**, 602-607.
- Harris, I., Jones, P.D., Osborn, T.J. & Lister, D.H. (2014) Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology*, **34**, 623-642.
- Hartkamp, A.D., De Beurs, K., Stein, A. & White, J.W. (1999) Interpolation Techniques for Climate Variables. In: *CIMMYT NRG-GIS Series*, pp. 1-34, Mexico City.
- Hartmann, D.L., Klein Tank, A.M.G., Rusticucci, M., Alexander, L.V., Brönnimann, S., Charabi, Y., Dentener, F.J., Dlugokencky, E.J., Easterling, D.R., Kaplan, A., Soden, B.J., Thorne, P.W., Wild, M. & Zhai, P.M. (2013) Observations: Atmosphere and Surface. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley), pp. 159–254. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Hawkins, D.M. & Cressie, N. (1984) Robust kriging—A proposal. *Journal of the International Association for Mathematical Geology*, **16**, 3-18.
- Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M. & Gräler, B. (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *Peerj*, **6**

- Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., de Jesus, J.M., Tamene, L. & Tondoh, J.E. (2015) Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *Plos One*, **10**
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. & Thépaut, J.-N. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999-2049.
- Hewitt, C.D., Allis, E., Mason, S.J., Muth, M., Pulwarty, R., Shumake-Guillemot, J., Bucher, A., Brunet, M., Fischer, A.M., Hama, A.M., Kolli, R.K., Lucio, F., Ndiaye, O. & Tapia, B. (2020) Making Society Climate Resilient: International Progress under the Global Framework for Climate Services. *Bulletin of the American Meteorological Society*, **101**, E237-E252.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.
- Hoaglin, D.C., Mosteller, F. & Tukey, J.W. (2000) *Understanding robust and exploratory data analysis*. John Wiley & Sons.
- Hoerling, M., Hurrell, J., Eischeid, J. & Phillips, A. (2006) Detection and Attribution of Twentieth-Century Northern and Southern African Rainfall Change. *Journal of Climate*, **19**, 3989-4008.
- Hoerling, M., Hurrell, J., Kumar, A., Terray, L., Eischeid, J., Pegion, P., Zhang, T., Quan, X. & Xu, T. (2011) On North American Decadal Climate for 2011–20. *Journal of Climate*, **24**, 4519-4528.
- Houle, D., Bouffard, A., Duchesne, L., Logan, T. & Harvey, R. (2012) Projections of Future Soil Temperature and Water Content for Three Southern Quebec Forested Sites. *Journal of Climate*, **25**, 7690-7701.
- Hu, X., Hou, Y., Li, D., Hua, T., Marchi, M., Paola Forero Urrego, J., Huang, B., Zhao, W. & Cherubini, F. (2023) Changes in multiple ecosystem services and their influencing factors in Nordic countries. *Ecological Indicators*, **146**, 109847.
- Hutchinson, M.F. (1995a) Interpolating mean rainfall using thin plate smoothing splines. *International journal of geographical information systems*, **9**, 385-403.
- Hutchinson, M.F. & Gessler, P.E. (1994) Splines — more than just a smooth interpolator. *Geoderma*, **62**, 45-67.
- Hutchinson, M.F. & Xu, T. (2013) ANUSPLIN Version 4.4 User Guide. In. Australian National University - Fenner School of Environment & Society
- Illés, G. & Móricz, N. (2022) Climate envelope analyses suggests significant rearrangements in the distribution ranges of Central European tree species. *Annals of Forest Science*, **79**, 35.
- IPCC (2021) *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom; New York, NY, USA.

- Iturbide, M., Gutiérrez, J.M., Alves, L.M., Bedia, J., Cerezo-Mota, R., Gimenez, E., Cofiño, A.S., Di Luca, A., Faria, S.H., Gorodetskaya, I.V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H.T., Jones, R.G., Krakovska, S., Manzanar, R., Martínez-Castro, D., Narisma, G.T., Nurhati, I.S., Pinto, I., Seneviratne, S.I., van den Hurk, B. & Vera, C.S. (2020) An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets. *Earth Syst. Sci. Data*, **12**, 2959-2970.
- Jarnevich, C.S., Young, N.E., Talbert, M. & Talbert, C. (2018) Forecasting an invasive species' distribution with global distribution data, local data, and physiological information. *Ecosphere*, **9**, e02279.
- Jiang, Y. (2009) Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models. *Energy*, **34**, 1276-1283.
- Jones, P.D., New, M., Parker, D.E., Martin, S. & Rigor, I.G. (1999) Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, **37**, 173-199.
- Karger, D.N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P. & Kessler, M. (2017) Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, **4**, 170122.
- Karmalkar, A.V., Thibeault, J.M., Bryan, A.M. & Seth, A. (2019) Identifying credible and diverse GCMs for regional climate change studies—case study: Northeastern United States. *Climatic Change*, **154**, 367-386.
- Katsavounidis, I., Kuo, C.C.J. & Zhang, Z. (1994) A new initialization technique for generalized Lloyd iteration. *IEEE Signal Processing Letters*, **1**, 144-146.
- Kelley, M., Schmidt, G.A., Nazarenko, L.S., Bauer, S.E., Ruedy, R., Russell, G.L., Ackerman, A.S., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Cesana, G., Cheng, Y., Clune, T.L., Cook, B.I., Cruz, C.A., Del Genio, A.D., Elsaesser, G.S., Faluvegi, G., Kiang, N.Y., Kim, D., Lacis, A.A., Leboissetier, A., LeGrande, A.N., Lo, K.K., Marshall, J., Matthews, E.E., McDermid, S., Mezuman, K., Miller, R.L., Murray, L.T., Oinas, V., Orbe, C., García-Pando, C.P., Perlwitz, J.P., Puma, M.J., Rind, D., Romanou, A., Shindell, D.T., Sun, S., Tausnev, N., Tsigaridis, K., Tselioudis, G., Weng, E., Wu, J. & Yao, M.-S. (2020) GISS-E2.1: Configurations and Climatology. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS002025.
- Kirtman, B., Power, S.B., Adedoyin, J.A., Boer, G.J., Bojariu, R., Camilloni, I., Doblas-Reyes, F.J., Fiore, A.M., Kimoto, M., Meehl, G.A., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G.J., Vecchi, G. & Wang, H.J. (2013) Near-term Climate Change: Projections and Predictability. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley), pp. 953–1028. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, G.A. (2010) Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate*, **23**, 2739-2758.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. (2006) World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, **15**, {259-263}.
- Lawrimore, J.H., Menne, M.J., Gleason, B.E., Williams, C.N., Wuertz, D.B., Vose, R.S. & Rennie, J. (2011) An overview of the Global Historical Climatology Network monthly

- mean temperature data set, version 3. *Journal of Geophysical Research-Atmospheres*, **116**
- Lee, J.-K. & Kim, Y.-O. (2017) Selection of representative GCM scenarios preserving uncertainties. *Journal of Water and Climate Change*, **8**, 641-651.
- Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J.P., Engelbrecht, F., Fischer, E., Fyfe, J.C., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S. & Zhou, T. (2021) Future Global Climate: Scenario-based Projections and Near-term Information. *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by C. Intergovernmental Panel on Climate), pp. 553-672. Cambridge University Press, Cambridge.
- Lee, T.C.K., Zwiers, F.W., Zhang, X. & Tsao, M. (2006) Evidence of Decadal Climate Prediction Skill Resulting from Changes in Anthropogenic Forcing. *Journal of Climate*, **19**, 5305-5318.
- Li, J. & Heap, A.D. (2008) *A review of spatial interpolation methods for environmental scientists*. Geoscience Australia, Australian Government Canberra, Australia.
- Li, J. & Heap, A.D. (2014) Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, **53**, 173-189.
- Liang, Y.X., Gillett, N.P. & Monahan, A.H. (2020) Climate Model Projections of 21st Century Global Warming Constrained Using the Observed Warming Trend. *Geophysical Research Letters*, **47**
- Lovell, J.T., MacQueen, A.H., Mamidi, S., Bonnette, J., Jenkins, J., Napier, J.D., Sreedasyam, A., Healey, A., Session, A., Shu, S., Barry, K., Bonos, S., Boston, L., Daum, C., Deshpande, S., Ewing, A., Grabowski, P.P., Haque, T., Harrison, M., Jiang, J., Kudrna, D., Lipzen, A., Pendergast, T.H., Plott, C., Qi, P., Saski, C.A., Shakirov, E.V., Sims, D., Sharma, M., Sharma, R., Stewart, A., Singan, V.R., Tang, Y., Thibivillier, S., Webber, J., Weng, X., Williams, M., Wu, G.A., Yoshinaga, Y., Zane, M., Zhang, L., Zhang, J., Behrman, K.D., Boe, A.R., Fay, P.A., Fritschi, F.B., Jastrow, J.D., Lloyd-Reilly, J., Martínez-Reyna, J.M., Matamala, R., Mitchell, R.B., Rouquette, F.M., Ronald, P., Saha, M., Tobias, C.M., Udvardi, M., Wing, R.A., Wu, Y., Bartley, L.E., Casler, M., Devos, K.M., Lowry, D.B., Rokhsar, D.S., Grimwood, J., Juenger, T.E. & Schmutz, J. (2021) Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature*, **590**, 438-444.
- Mahony, C.R., Wang, T.L., Hamann, A. & Cannon, A.J. (2022) A global climate model ensemble for downscaled monthly climate normals over North America. *International Journal of Climatology*, **42**, 5871-5891.
- Maidment, R.I., Grimes, D., Allan, R.P., Tarnavsky, E., Stringer, M., Hewison, T., Roebeling, R. & Black, E. (2014) The 30 year TAMSAT African Rainfall Climatology And Time series (TARCAT) data set. *Journal of Geophysical Research: Atmospheres*, **119**, 10,619-10,644.
- Maier, H.R. & Dandy, G.C. (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, **15**, 101-124.
- Mandler, H. & Weigand, B. (2023) Feature importance in neural networks as a means of interpretation for data-driven turbulence models. *Computers & Fluids*, **265**, 105993.

- Marchi, M., Sinjur, I., Bozzano, M. & Westergren, M. (2019) Evaluating WorldClim Version 1 (1961–1990) as the Baseline for Sustainable Use of Forest and Environmental Resources in a Changing Climate. In: *Sustainability*
- Marchi, M., Castellanos-Acuña, D., Hamann, A., Wang, T., Ray, D. & Menzel, A. (2020) ClimateEU, scale-free climate normals, historical time series, and future projections for Europe. *Scientific Data*, **7**, 428.
- Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K. & Lonnoy, E. (2021) 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. *IPCC* (ed. by J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou), pp. 2087-2138. Cambridge University Press.
- Masson, D. & Knutti, R. (2011) Climate model genealogy. *Geophysical Research Letters*, **38**
- Matheron, G. (1963) Principles of geostatistics. *Economic Geology*, **58**, 1246-1266.
- Matheron, G. (1971) *The Theory of Regionalized Variables and Its Applications*. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Paris, France.
- Mathur, M., Mathur, P. & Purohit, H. (2023) Ecological niche modelling of a critically endangered species *Commiphora wightii* (Arn.) Bhandari using bioclimatic and non-bioclimatic variables. *Ecological Processes*, **12**, 8.
- Mbogga, M.S., Hansen, S., Wang, T. & Hamann, A. (2010) *A comprehensive set of interpolated climate data for Alberta*. . Government of Alberta, Alberta Sustainable Resource Development.
- McCulloch, W.S. & Pitts, W. (1990) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, **52**, 99-115.
- McSweeney, C.F., Jones, R.G., Lee, R.W. & Rowell, D.P. (2015) Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dynamics*, **44**, 3237-3260.
- Meehl, G.A., Senior, C.A., Eyring, V., Flato, G., Lamarque, J.F., Stouffer, R.J., Taylor, K.E. & Schlund, M. (2020) Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, **6**
- Meinshausen, M., Nicholls, Z.R.J., Lewis, J., Gidden, M.J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J.G., Daniel, J.S., John, A., Krummel, P.B., Luderer, G., Meinshausen, N., Montzka, S.A., Rayner, P.J., Reimann, S., Smith, S.J., van den Berg, M., Velders, G.J.M., Vollmer, M.K. & Wang, R.H.J. (2020) The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geosci. Model Dev.*, **13**, 3571-3605.
- Mejia, J.M. & Rousselle, J. (1976) Disaggregation models in hydrology revisited. *Water Resources Research*, **12**, 185-186.
- Mekonnen, Z.A., Riley, W.J., Randerson, J.T., Grant, R.F. & Rogers, B.M. (2019) Expansion of high-latitude deciduous forests driven by interactions between climate warming and fire. *Nature plants*, **5**, 952-958.
- Mendlik, T. & Gobiet, A. (2016) Selecting climate simulations for impact studies based on multivariate patterns of climate change. *Climatic Change*, **135**, 381-393.
- Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E. & Houston, T.G. (2012) An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*, **29**, 897-910.

- Minaei, A., Todeschini, S., Sitzenfrei, R. & Creaco, E. (2022) Ensemble Evaluation and Member Selection of Regional Climate Models for Impact Models Assessment. In: *Water*
- Mote, P.W., Allen, M.R., Jones, R.G., Li, S.H., Mera, R., Rupp, D.E., Salahuddin, A. & Vickers, D. (2016) Superensemble Regional Climate Modeling for the Western United States. *Bulletin of the American Meteorological Society*, **97**, 203-216.
- Müller, W.A., Jungclaus, J.H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornbluh, L., Li, H., Modali, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., Tian, F. & Marotzke, J. (2018) A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, **10**, 1383-1413.
- Nalder, I.A. & Wein, R.W. (1998) Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest. *Agricultural and Forest Meteorology*, **92**, 211-225.
- New, M., Hulme, M. & Jones, P. (1999) Representing Twentieth-Century space-time climate variability. Part I: Development of a 1961-90 mean monthly terrestrial climatology. *Journal of Climate*, **12**, 829-856.
- Ngarega, B.K., Chaibva, P., Masocha, V.F., Saina, J.K., Khine, P.K. & Schneider, H. (2023) Application of MaxEnt modeling to evaluate the climate change effects on the geographic distribution of *Lippia javanica* (Burm.f.) Spreng in Africa. *Environmental Monitoring and Assessment*, **196**, 62.
- Nicholson, S.E. (2013) The West African Sahel: A Review of Recent Studies on the Rainfall Regime and Its Interannual Variability. *ISRN Meteorology*, **2013**, 453521.
- NOAA (2018) Federal climate complex data documentation for integrated surface data (ISD). In: NOAA - National Centers for Environmental Information, Asheville, NC, USA.
- Novella, N.S. & Thiaw, W.M. (2013) African Rainfall Climatology Version 2 for Famine Early Warning Systems. *Journal of Applied Meteorology and Climatology*, **52**, 588-606.
- Nychka, D., Furrer, R., Paige, J. & Sain, S. (2021) fields: Tools for spatial data. In: *R package version 15.2*. University Corporation for Atmospheric Research, Boulder, CO, USA.
- Ogougbé, R.E., Agbo, R.I., Ahamidé, B., Djego-Djossou, S. & Djego, G.J. (2022) Potential geographic distribution and modelling of the ecological niche of *Harrisonia abyssinica*, a priority medicinal tree species in Benin (West Africa). *Modeling Earth Systems and Environment*, **8**, 2471-2483.
- Omukuti, J., Wanzala, M.A., Ngaina, J. & Ganola, P. (2023) Develop medium- to long-term climate information services to enhance comprehensive climate risk management in Africa. *Climate Resilience and Sustainability*, **2**
- Overland, J.E., Wang, M.Y., Walsh, J.E. & Stroeve, J.C. (2014) Future Arctic climate changes: Adaptation and mitigation time scales. *Earths Future*, **2**, 68-74.
- Pang, B., Yue, J., Zhao, G. & Xu, Z. (2017) Statistical Downscaling of Temperature with the Random Forest Model. *Advances in Meteorology*, **2017**, 7265178.
- Parisien, M.-A., Barber, Q.E., Bourbonnais, M.L., Daniels, L.D., Flannigan, M.D., Gray, R.W., Hoffman, K.M., Jain, P., Stephens, S.L. & Taylor, S.W. (2023) Abrupt, climate-induced increase in wildfires in British Columbia since the mid-2000s. *Communications Earth & Environment*, **4**, 309.
- Parks, S.A., Holsinger, L.M., Panunto, M.H., Jolly, W.M., Dobrowski, S.Z. & Dillon, G.K. (2018) High-severity fire: evaluating its key drivers and mapping its probability across western US forests. *Environmental research letters*, **13**, 044037.

- PCMDI, P.f.C.M.D.I. (2022) *The Program for Climate Model Diagnosis and Intercomparison: furthering the understanding of climate change through model and observational analysis and community leadership*. Available at: <https://pcmdi.llnl.gov/about.html> (accessed April, 20 2024).
- PCMDI, P.f.C.M.D.I. (2024) *ESGF CMIP6 Data Holdings*. Available at: https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf_data_holdings/index.html (accessed April, 20 2024).
- Räsänen, J. & Ruokolainen, L. (2006) Probabilistic forecasts of near-term climate change based on a resampling ensemble technique. *Tellus A: Dynamic Meteorology and Oceanography*, **58**, 461-472.
- Rampal, N., Gibson, P.B., Sood, A., Stuart, S., Fauchereau, N.C., Brandolino, C., Noll, B. & Meyers, T. (2022) High-resolution downscaling with interpretable deep learning: Rainfall extremes over New Zealand. *Weather and Climate Extremes*, **38**, 100525.
- Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fife, J., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A. & Taylor, K.E. (2007) Climate Models and Their Evaluation. *The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Rigol, J.P., Jarvis, C.H. & Stuart, N. (2001) Artificial neural networks as a tool for spatial interpolation. *International Journal of Geographical Information Science*, **15**, 323-343.
- Rubel, F. & Kottek, M. (2010) Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift*, **19**, {135-141}.
- Sáenz-Romero, C., O'Neill, G., Aitken, S.N. & Lindig-Cisneros, R. (2020) Assisted migration field tests in Canada and Mexico: Lessons, limitations, and challenges. *Forests*, **12**, 9.
- Schirpke, U. & Ebner, M. (2022) Exposure to global change pressures and potential impacts on ecosystem services of mountain lakes in the European Alps. *Journal of Environmental Management*, **318**, 115606.
- Schlenker, W., Hanemann, W.M. & Fisher, A.C. (2007) Water Availability, Degree Days, and the Potential Impact of Climate Change on Irrigated Agriculture in California. *Climatic Change*, **81**, 19-38.
- Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voltaire, A., Colin, J., Decharme, B., Delire, C., Berthet, S., Chevallier, M., Sénési, S., Franchisteguy, L., Vial, J., Mallet, M., Joetzjer, E., Geoffroy, O., Guérémy, J.-F., Moine, M.-P., Msadek, R., Ribes, A., Rocher, M., Roehrig, R., Salas-y-Méla, D., Sanchez, E., Terray, L., Valcke, S., Waldman, R., Aumont, O., Bopp, L., Deshayes, J., Éthé, C. & Madec, G. (2019) Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate. *Journal of Advances in Modeling Earth Systems*, **11**, 4182-4227.
- Sekulić, A., Kilibarda, M., Heuvelink, G.B.M., Nikolić, M. & Bajat, B. (2020) Random Forest Spatial Interpolation. In: *Remote Sensing*
- Sellar, A.A., Jones, C.G., Mulcahy, J.P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F.M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S.T., Kelley, D.I., Ellis, R., Johnson, C.E., Walton, J., Abraham, N.L., Andrews, M.B., Andrews, T., Archibald, A.T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi,

- M., Edwards, J., Folberth, G.A., Gedney, N., Griffiths, P.T., Harper, A.B., Hendry, M.A., Hewitt, A.J., Johnson, B., Jones, A., Jones, C.D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R.J., Predoi, V., Robertson, E., Siahann, A., Smith, R.S., Swaminathan, R., Woodhouse, M.T., Zeng, G. & Zerroukat, M. (2019) UKESM1: Description and Evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*, **11**, 4513-4558.
- Seo, S.B., Kim, Y.O., Kim, Y. & Eum, H.I. (2019) Selecting climate change scenarios for regional hydrologic impact studies based on climate extremes indices. *Climate Dynamics*, **52**, 1595-1611.
- Shepard, D. (1968) A two-dimensional interpolation function for irregularly-spaced data. In: *Proceedings of the 1968 23rd ACM national conference*, pp. 517–524. Association for Computing Machinery
- Shrestha, A.B., Wake, C.P., Mayewski, P.A. & Dibb, J.E. (1999) Maximum Temperature Trends in the Himalaya and Its Vicinity: An Analysis Based on Temperature Records from Nepal for the Period 1971–94. *Journal of Climate*, **12**, 2775-2786.
- Smith, K. (1975) *Principles of Applied Climatology*. Wiley.
- Snell, S.E., Gopal, S. & Kaufmann, R.K. (2000) Spatial Interpolation of Surface Air Temperatures Using Artificial Neural Networks: Evaluating Their Use for Downscaling GCMs. *Journal of Climate*, **13**, 886-895.
- Sung, J.H., Kwon, M., Jeon, J.-J. & Seo, S.B. (2019) A Projection of Extreme Precipitation Based on a Selection of CMIP5 GCMs over North Korea. In: *Sustainability*
- Suzuki, R., Terada, Y. & Shimodaira, H. (2019) *pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling*.
- 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., Sigmund, M., Solheim, L., von Salzen, K., Yang, D. & Winter, B. (2019) The Canadian Earth System Model version 5 (CanESM5.0.3). *Geosci. Model Dev.*, **12**, 4823-4873.
- Tan, J., Xie, X., Zuo, J., Xing, X., Liu, B., Xia, Q. & Zhang, Y. (2021) Coupling random forest and inverse distance weighting to generate climate surfaces of precipitation and temperature with Multiple-Covariates. *Journal of Hydrology*, **598**, 126270.
- Tank, A.M.G.K., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., Van Engelen, A.F.V., Forland, E., Miletus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., López, J.A., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L.V. & Petrovic, P. (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology*, **22**, 1441-1453.
- Tarboton, D.G., Sharma, A. & Lall, U. (1998) Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water Resources Research*, **34**, 107-119.
- 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. & Kimoto, M. (2019) Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.*, **12**, 2727-2765.
- Taylor, K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, **106**, 7183-7192.
- Tett, S.F.B., Stott, P.A., Allen, M.R., Ingram, W.J. & Mitchell, J.F.B. (1999) Causes of twentieth-century temperature change near the Earth's surface. *Nature*, **399**, 569-572.
- Thyer, M. & Kuczera, G. (2000) Modeling long-term persistence in hydroclimatic time series using a hidden state Markov Model. *Water Resources Research*, **36**, 3301-3310.
- USGS, U.S.G.S. (1996) GTOPO30, A global digital elevation model (DEM). In:
- van Niekerk, A. & Joubert, S.J. (2011) Input variable selection for interpolating high-resolution climate surfaces for the Western Cape. *Water SA*, **37**, 271-279.
- Verdin, A., Funk, C., Peterson, P., Landsfeld, M., Tuholske, C. & Grace, K. (2020) Development and validation of the CHIRTS-daily quasi-global high-resolution daily temperature data set. *Scientific Data*, **7**, 303.
- Volodin, E.M., Mortikov, E.V., Kostykin, S.V., Galin, V.Y., Lykossov, V.N., Gritsun, A.S., Diansky, N.A., Gusev, A.V. & Iakovlev, N.G. (2017) Simulation of the present-day climate with the climate model INMCM5. *Climate Dynamics*, **49**, 3715-3734.
- Vose, R.S., McNeill, S., Thomas, K. & Shepherd, E. (2011) Enhanced Master Station History Report of March 2019. In. NOAA National Climatic Data Center
- Wahba, G. (1979) How to Smooth Curves and Surfaces with Splines and Cross-Validation. In. Wisconsin University - Madison Department of Statistics
- Wang, T., Hamann, A., Spittlehouse, D.L. & Aitken, S.N. (2006) Development of scale-free climate data for western Canada for use in resource management. *Int. J. Climatol.*, **26**, 383-397.
- Wang, T., Hamann, A., Spittlehouse, D.L. & Murdock, T.Q. (2012) ClimateWNA—High-Resolution Spatial Climate Data for Western North America. *Journal of Applied Meteorology and Climatology*, **51**, 16-29.
- Wang, T.L., Hamann, A., Spittlehouse, D. & Carroll, C. (2016) Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America. *Plos One*, **11**
- Weigel, A.P., Knutti, R., Liniger, M.A. & Appenzeller, C. (2010) Risks of Model Weighting in Multimodel Climate Projections. *Journal of Climate*, **23**, 4175-4191.
- Wesselkamp, M., Roberts, D.R. & Dormann, C.F. (2024) Identifying potential provenances for climate-change adaptation using spatially variable coefficient models. *BMC Ecology and Evolution*, **24**, 70.
- Whittleston, D., Nicholson, S.E., Schlosser, A. & Entekhabi, D. (2017) Climate Models Lack Jet-Rainfall Coupling over West Africa. *Journal of Climate*, **30**, 4625-4632.
- Wilcke, R.A.I. & Barring, L. (2016) Selecting regional climate scenarios for impact modelling studies. *Environmental Modelling & Software*, **78**, 191-201.
- Wilks, D.S. (2019) Chapter 10 - Time Series. *Statistical Methods in the Atmospheric Sciences (Fourth Edition)* (ed. by D.S. Wilks), pp. 485-550. Elsevier.
- Willmott, C.J. & Matsuura, K. (1995) Smart Interpolation of Annually Averaged Air Temperature in the United States. *Journal of Applied Meteorology (1988-2005)*, **34**, 2577-2586.

- WMO (1996) Climatological Normals (CLINO) for the period 1961-1990. In, p. 788. World Meteorological Organization, Geneva, Switzerland.
- WMO, W.M.O. (2022) *Essential Climate Variables*. Available at: <https://public.wmo.int/en/programmes/global-climate-observing-system/essential-climate-variables> (accessed April 01 2023)
- Wolfensberger, D., Gabella, M., Boscacci, M., Germann, U. & Berne, A. (2021) RainForest: a random forest algorithm for quantitative precipitation estimation over Switzerland. *Atmospheric Measurement Techniques*, **14**, 3169-3193.
- Workneh, H.T., Chen, X., Ma, Y., Bayable, E. & Dash, A. (2024) Comparison of IDW, Kriging and orographic based linear interpolations of rainfall in six rainfall regimes of Ethiopia. *Journal of Hydrology: Regional Studies*, **52**, 101696.
- Wu, T. & Li, Y. (2013) Spatial interpolation of temperature in the United States using residual kriging. *Applied Geography*, **44**, 112-120.
- Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., Jie, W., Zhang, J., Liu, Y., Zhang, L., Zhang, F., Zhang, Y., Wu, F., Li, J., Chu, M., Wang, Z., Shi, X., Liu, X., Wei, M., Huang, A., Zhang, Y. & Liu, X. (2019) The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6. *Geosci. Model Dev.*, **12**, 1573-1600.
- Xie, P. & Arkin, P.A. (1997) Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs. *Bulletin of the American Meteorological Society*, **78**, 2539-2558.
- Yatagai, A., Arakawa, O., Kamiguchi, K., Kawamoto, H., Nodzu, M.I. & Hamada, A. (2010) A 44-year daily gridded precipitation dataset for Asia based on a dense network of rain gauges, SOLA. In:
- Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N. & Kitoh, A. (2012) APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges. *Bulletin of the American Meteorological Society*, **93**, 1401-1415.
- Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y. & Ishii, M. (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. Ser. II*, **advpub**
- Zeng, Z., Wang, Z., Gui, K., Yan, X., Gao, M., Luo, M., Geng, H., Liao, T., Li, X., An, J., Liu, H., He, C., Ning, G. & Yang, Y. (2020) Daily Global Solar Radiation in China Estimated From High-Density Meteorological Observations: A Random Forest Model Framework. *Earth and Space Science*, **7**, e2019EA001058.
- Zhan, J., Wu, S., Qi, J., Zeng, J., Qin, M., Wang, Y. & Du, Z. (2023) A generalized spatial autoregressive neural network method for three-dimensional spatial interpolation. *Geosci. Model Dev.*, **16**, 2777-2794.
- Ziehn, T., Chamberlain, M.A., Law, R.M., Lenton, A., Bodman, R.W., Dix, M., Stevens, L., Wang, Y.-P. & Srbinovsky, J. (2020) The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, **70**, 193-214.

Appendix

A. Initial thin-plate spline interpolation with binned weather station data

```
### Training data with LAT LONG ELEV and climate VARIABLES
dat1 = read.csv("weather_station_data.csv")
head(dat1)
nrow(dat1)

### Aggregation by d degrees and m meter elevation bins
d = 3
m = 250
dat1$LAT1 = round(dat1$LAT/d)*d
dat1$LONG1 = round(dat1$LONG/d)*d
dat1$ELEV1 = round(dat1$ELEV/m)*m
data.frame(names(dat1))
dat2 = aggregate(dat1[,c("LONG", "LAT", "ELEV", "VAR")],
                 by=list(LAT1=dat1$LAT1, LONG1=dat1$LONG1,
                        ELEV1=dat1$ELEV1), mean, na.rm=T)
head(dat2)
nrow(dat2)

### Target grid in xyz triplet format with LAT LONG ELEV
dat3 = read.csv("target_grid_as_xyz.csv")

### Thin-plate spline interpolation
library(fields)
out1 = fastTps(dat2[,c("LONG", "LAT", "ELEV")], dat2$VAR,
              lon.lat=T, aRange=2000)

### TPS prediction on stations for neural network fine-tuning
dat1$VAR_tps = round(predict(out1, dat1[,c("LONG", "LAT", "ELEV")]), 1)
head(dat1)

### TPS prediction on target grid for visual checks
dat3$VAR_tps = round(predict(out1, dat3[,c("LONG", "LAT", "ELEV")]), 1)
head(dat3)

### Conversion to GIS-readable ascii grid
library(remotes)
install_github("cran/SDMTools") # requires RTools.exe install
library(SDMTools)
dataframe2asc(dat3[,c("LAT", "LONG", "VAR_tps")])
```

B. Fine-tuning the thin-plate spline interpolations with a neural network

```
### Requires separate installation of Anaconda with tensorflow-gpu.
library(keras)
use_condaenv('base', required=T)

### Input data are vector of the dependent variable VAR (y_train)
### and a matrix of scaled predictor variables (x_train), including
### the main variables VAR_tps, LAT, LONG, ELEV and all covariates

### Define neural network
model = keras_model_sequential() %>%
  layer_dense(units = 2048, activation = "relu", input_shape =
    ncol(x_train), kernel_regularizer = regularizer_l2(0.01)) %>%
  layer_dropout(rate = 0.2) %>%
  layer_dense(units = 1024, activation = "relu") %>%
  layer_dense(units = 512, activation = "relu") %>%
  layer_dense(units = 256, activation = "relu") %>%
  layer_dense(units = 128, activation = "relu") %>%
  layer_dense(units = 64, activation = "relu") %>%
  layer_dense(units = 32, activation = "relu") %>%
  layer_dense(units = 16, activation = "relu") %>%
  layer_dense(units = 1)

### Compile neural network
model %>% compile(
  optimizer = 'adam',
  loss = 'mean_squared_error',
  metrics = c('mean_squared_error')
)

### Train the model
history = model %>% fit(x_train, y_train,
  epochs = 150,
  batch_size = 32,
  # validation_split = 0.2, # only for development
  verbose = 1)

### Importance values
library(DALEX)
impl = variable_importance(explain(model, x_train, y_train))
plot(impl)
impl = data.frame(impl)
impl = aggregate(impl[,3], by=list(variable=impl$variable), mean)
write.csv(impl, 'Importance Values VAR.csv')

### Predictions on target grid, which includes the same, identically
### scaled predictor variables as in the x_train matrix
pred1 = model %>% predict(target)
```