A high-resolution database of historical and future climate for Africa developed with deep neural networks

Sarah Namiiro[1], Andreas Hamann[1]*, Tongli Wang[2], Dante Castellanos-Acuña[1], Colin Mahoney[3]

[1] Department of Renewable Resources, University of Alberta, 751 General Services Building, Edmonton, AB, Canada, T6G 2H1
[2] Centre for Forest Conservation Genetics, Department of Forest and Conservation Sciences, University of British Columbia.
[3] British Columbia Ministry of Forests, Victoria, BC, Canada.

(Manuscript under review)

* Corresponding author's contact information:
Tel: (780) 492-6429
Fax: (780) 492-4323
Email: andreas.hamann@ualberta.ca

Abstract: This study contributes an accessible, comprehensive database of interpolated climate data for Africa that includes monthly, annual, decadal, and 30-year normal climate data for the last 120 years (1901 to present) as well as multi-model CMIP6 climate change projections for the 21st century. The database includes variables relevant for ecological research and infrastructure planning, and comprises more than 25,000 climate grids that can be queried with a provided *ClimateAF* software package. In addition, 30 arcsecond (~1km) resolution gridded data are available for download. The climate grids were developed with a three-step approach, using thin-plate spline interpolations of weather station data as a first approximation. Subsequently, a novel deep learning approach is used to model orographic precipitation, rain shadows, lake and coastal effects at moderate resolution. Lastly, lapse-rate based downscaling is applied to generate high-resolution grids. The climate estimates were optimized and cross-validated with a checkerboard approach to ensure that training data was spatially distanced from validation data. We conclude with a discussion of applications and limitations of this database.

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 is the final number of stations that were retained after applying quality control criteria, duplicate station removal and local redundancy control.

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, et al. [3]
Global Historic Climate Network Monthly (GHCN-M) , Version 3	1878-2017	Monthly time series	864 (176)	Lawrimore, et al. [22]
Global Historic Climate Network Daily (GHCN-D), Version 3	1900-2021	Daily time series	878 (344)	Menne, et al. [23]
World-wide Agroclimatic Data of FAO (FAOCLIM), Version 2	1902-1998	Monthly time series	846 (96)	FAO [24]
World Meteorological Organization (WMO)	1961-1992	Monthly time series	431 (256)	WMO [25]
European Climate Assessment Dataset (ECA)	1892-2018	Daily & Monthly time series	223 (13)	Tank, et al. [26]
National Oceanic & Atmospheric Administration (NOAA)	1949-2015	Monthly time series	131	NOAA [27]
Global monthly weather station for precipitation	1901-2010	Monthly time series	4510	Castellanos-Acuna and Hamann [28]

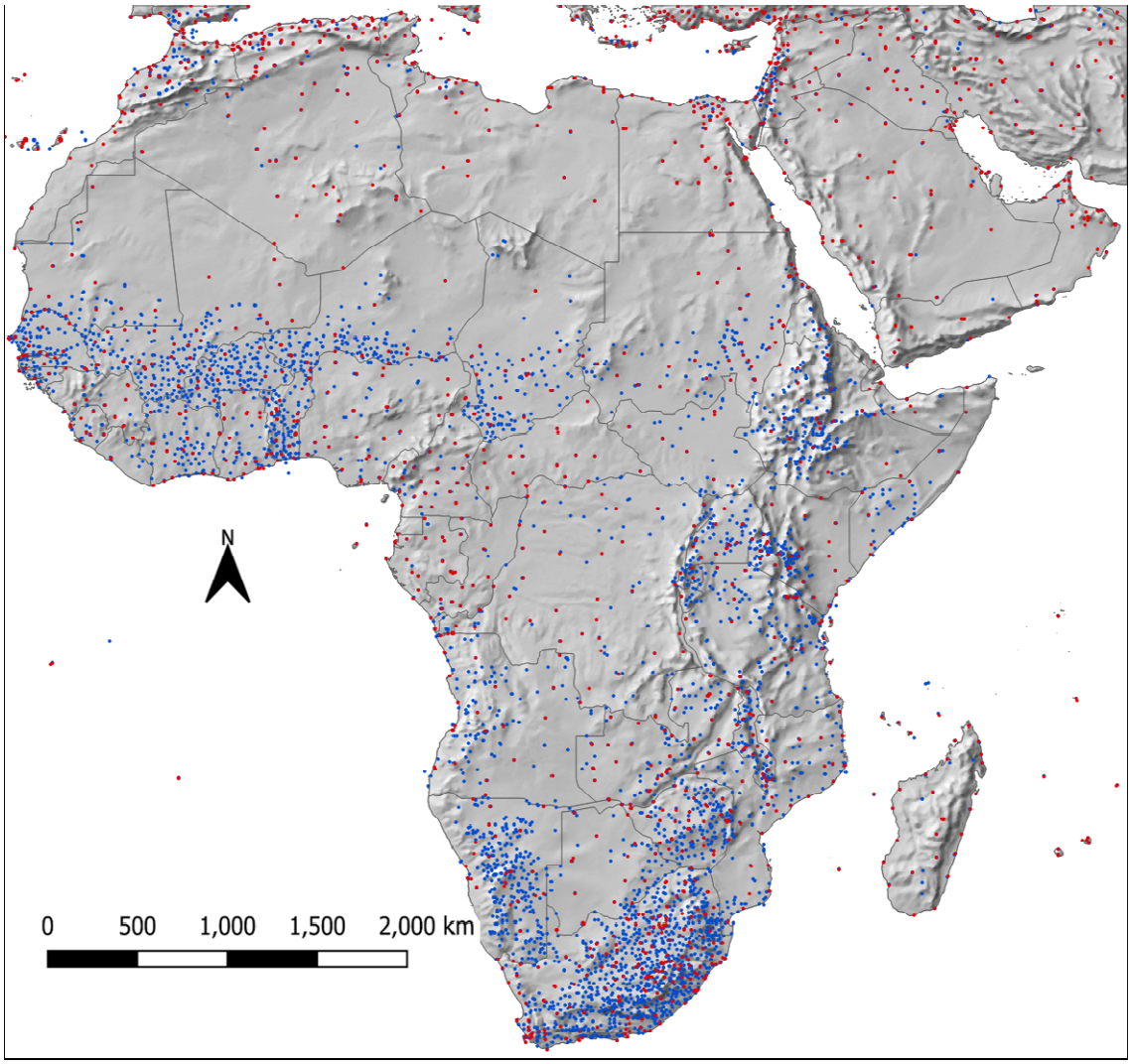


Figure 1: Distribution of 4625 weather stations compiled for the database. Blue stations have records for only precipitation measurements and red stations have records for both precipitation and temperature measurements for the 1961-1990 period.

Table 2. Predictor variables selected 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		
Base variables			
Thin-plate spline interpolation of climate variable			
Latitude			
Longitude			
Topographic variables			
Elevation	3	7	15
Compound topographic index	5	9	
Topographic position index	3	7	
Hill shade south-north direction	7		
Monthly variables weighted by wind direction and strength			
Windward (+) or leeward (–) slope exposure	5	9	15
Leeward wind-weighted distance to coast (max 50km)	5		
Leeward wind-weighted distance to coast (max 500km)	15		
Leeward wind-weighted distance to lakes (max 10km)	5		
Leeward wind-weighted distance to lakes (max 100km)	15		

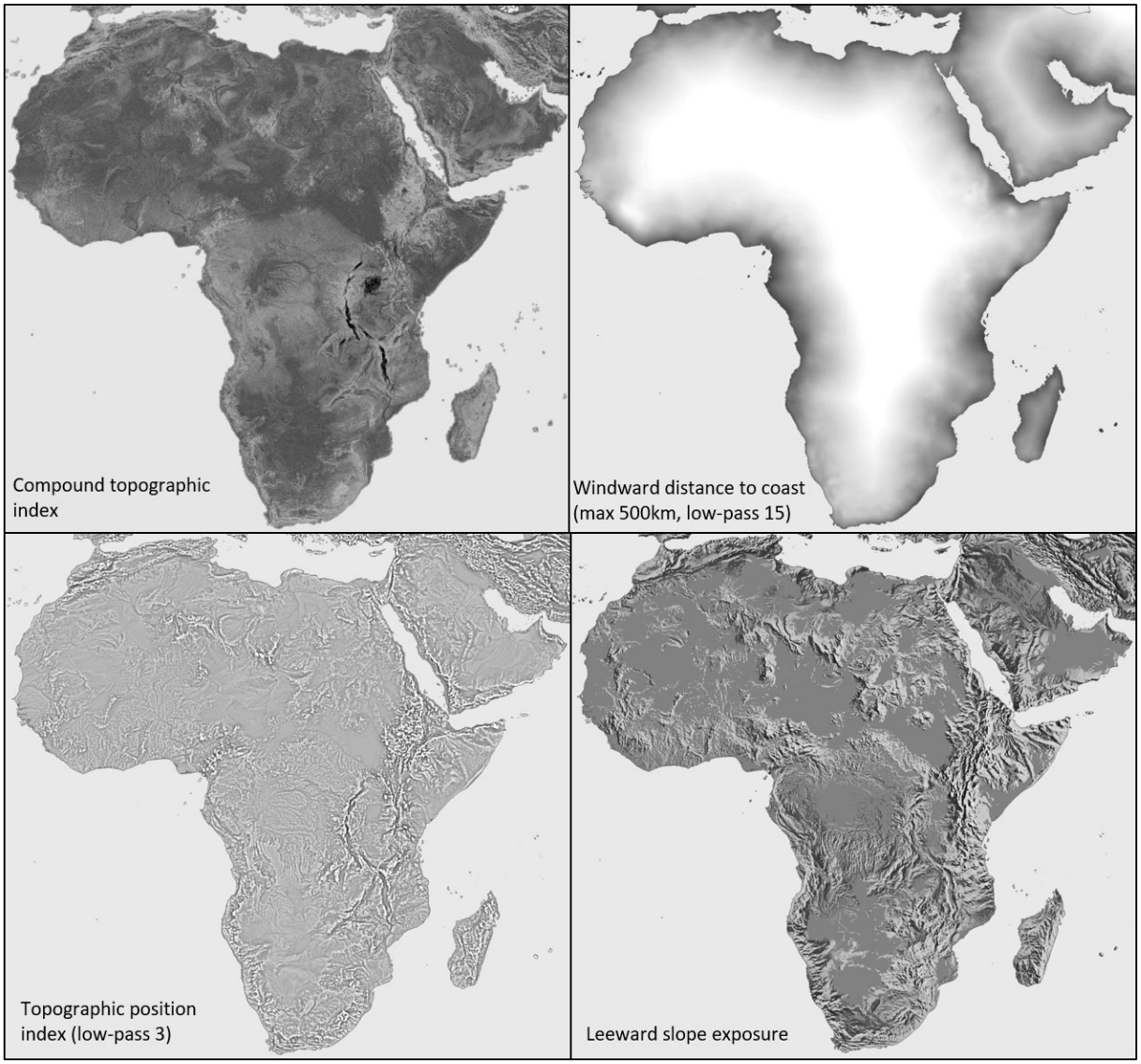


Figure 2: Example of predictor variables used in neural network fine-tuning of thin-plate spline interpolations. All putative predictor variables were subjected to transformations for normality if possible, and then scaled to values between 0 (black) and 1 (white) for use as covariates in neural network models.

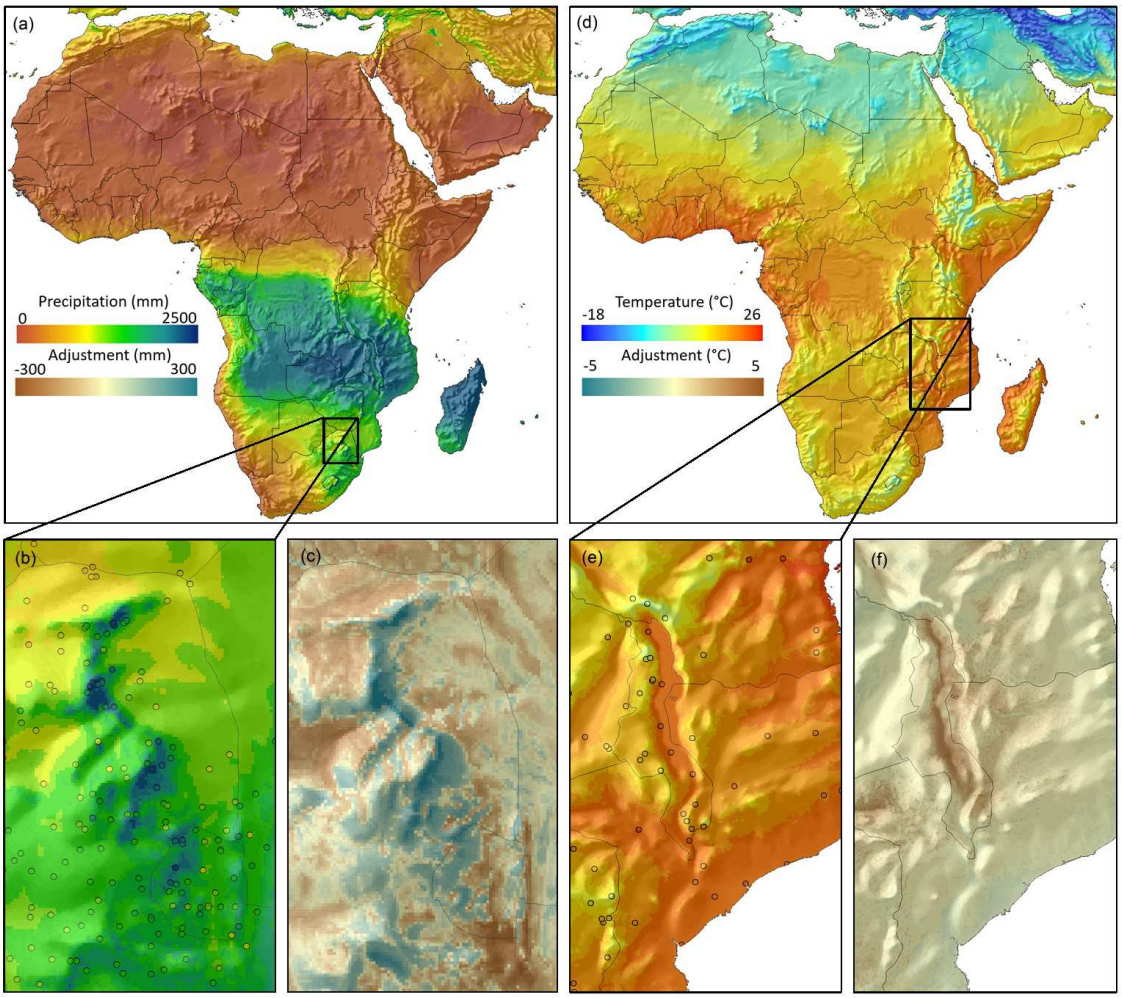


Figure 3. Example climate grid for January precipitation (**a-c**) and January mean minimum temperature (**d-f**), including a difference calculations (**c**, **f**) that highlight the effects of neural network fine-tuning. The color of circles in the inset indicates the weather station values (or residuals on the difference layer) on the same scale as the gridded data.

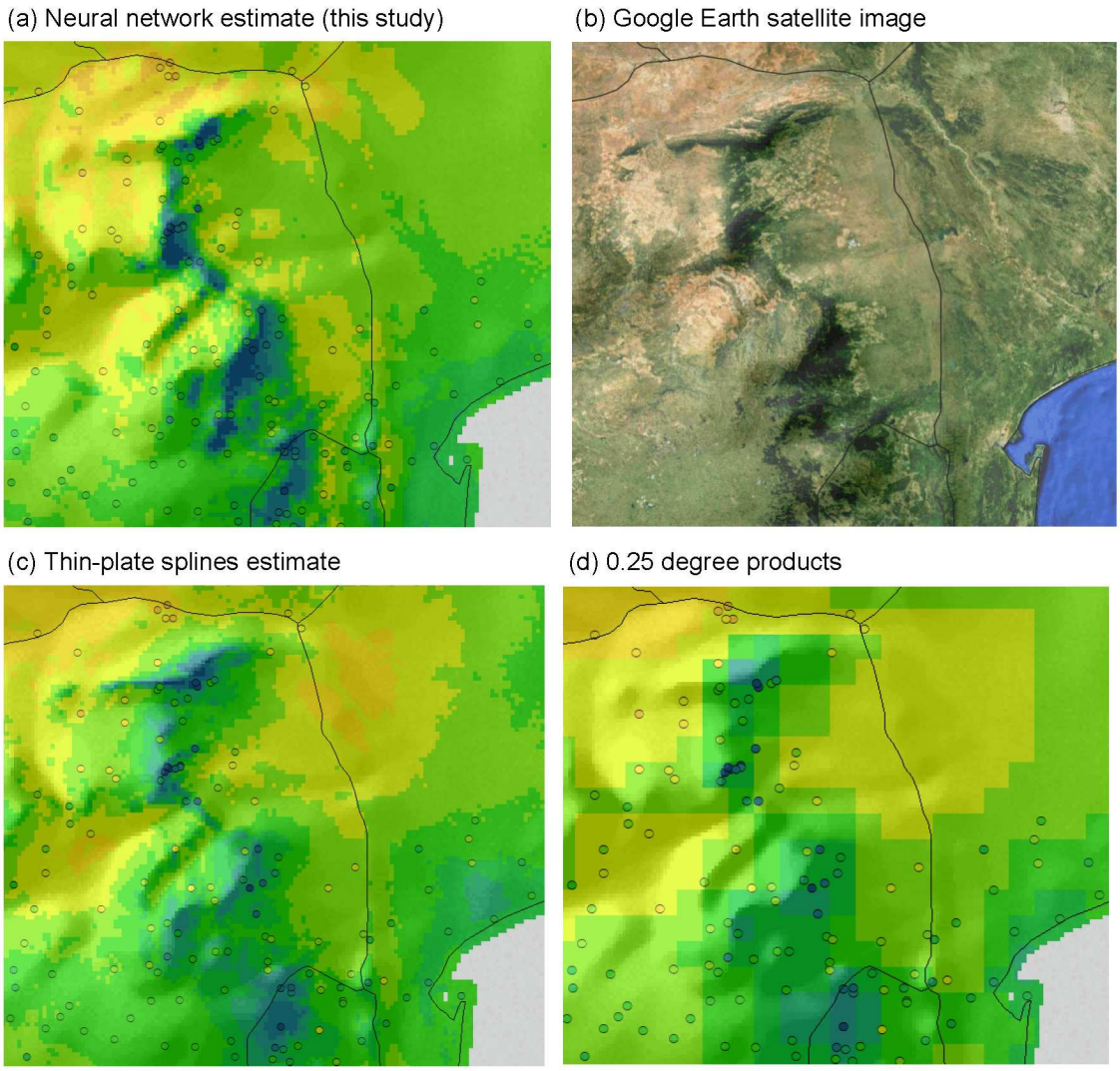
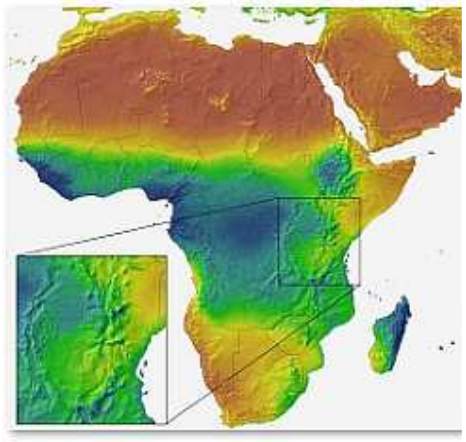


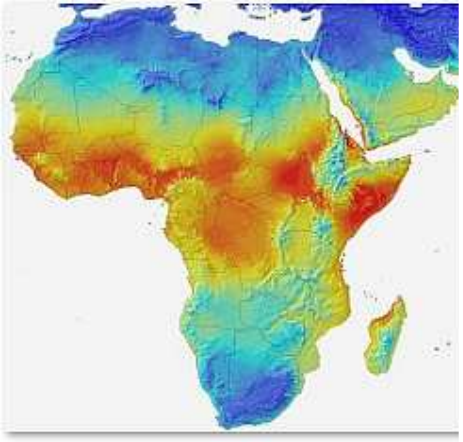
Figure 4. Comparison of interpolated climate grids for January precipitation, fine-tuned with neural networks from this study (a), with a Google Earth satellite image (b), the widely used thin-plate spline interpolations (c), and a commonly used grid size for time series data (d).

Figure 5: Sample grids for visual inspection of different variables. Follow the View link for 2.5 arcminute resolution JPGs with hillshades added for orientation. Once the image opens in a browser window, you can click on them to zoom in or out of different areas (🔍). The inset shows detail for major mountain ranges (Rift Valley, Mt Kenya, Mt Kilimanjaro area). To view the grids in GIS, use the RGB-GeoTIFFs which do not include hillshading (Note, corresponding data files are available at the bottome of this page http://tinyurl.com/ClimateAF).

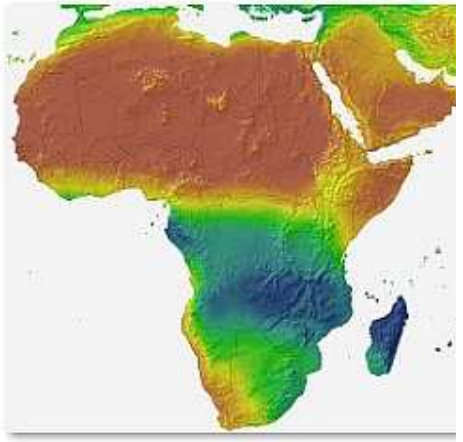
Mean Annual Precipitation (View, GIS)



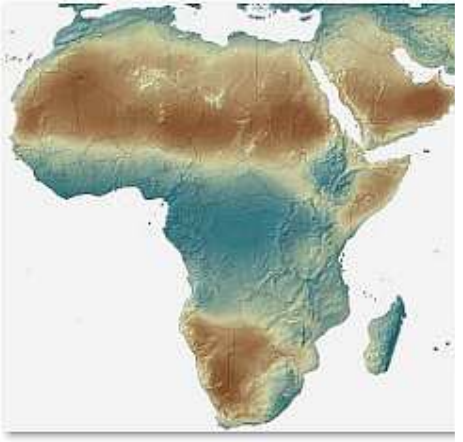
Mean Coldest Month Temp (View, GIS)



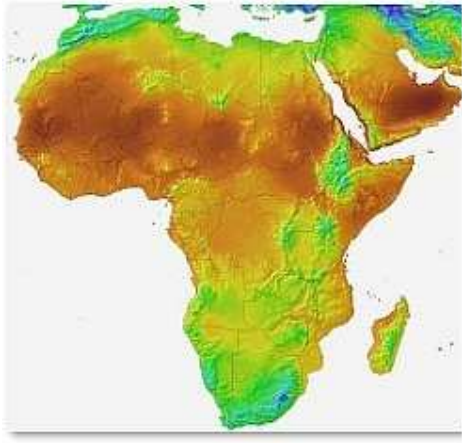
Precipitation Dec-Jan-Feb (View, GIS)



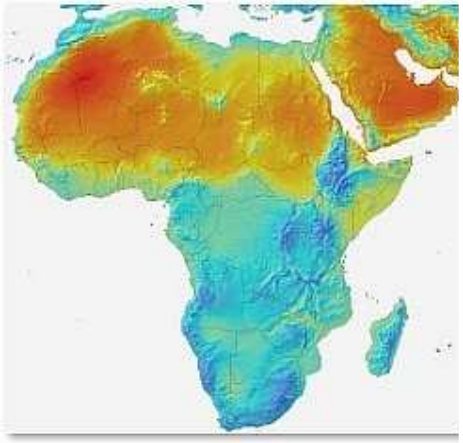
Climate Moisture Deficit (View, GIS)



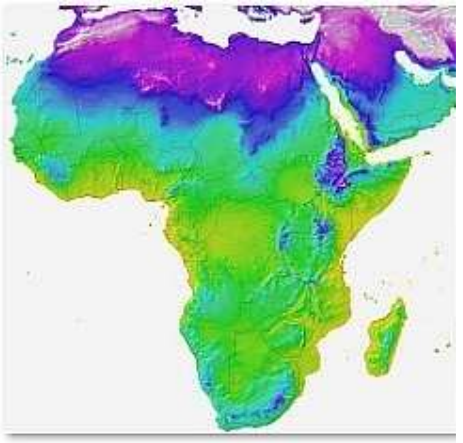
Mean Annual Temperature (View, GIS)



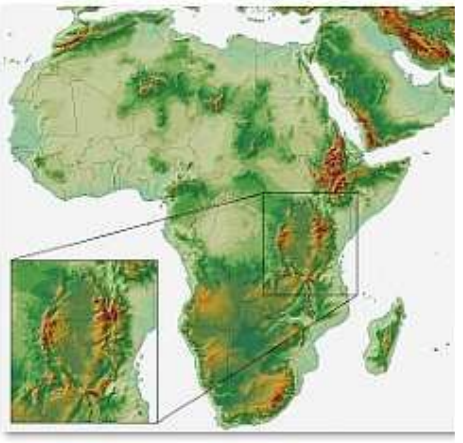
Mean Warmest Month Temp (View, GIS)



Avg Min Temp Dec-Jan-Feb (View, GIS)



Reference Elevation Grid (View, GIS)



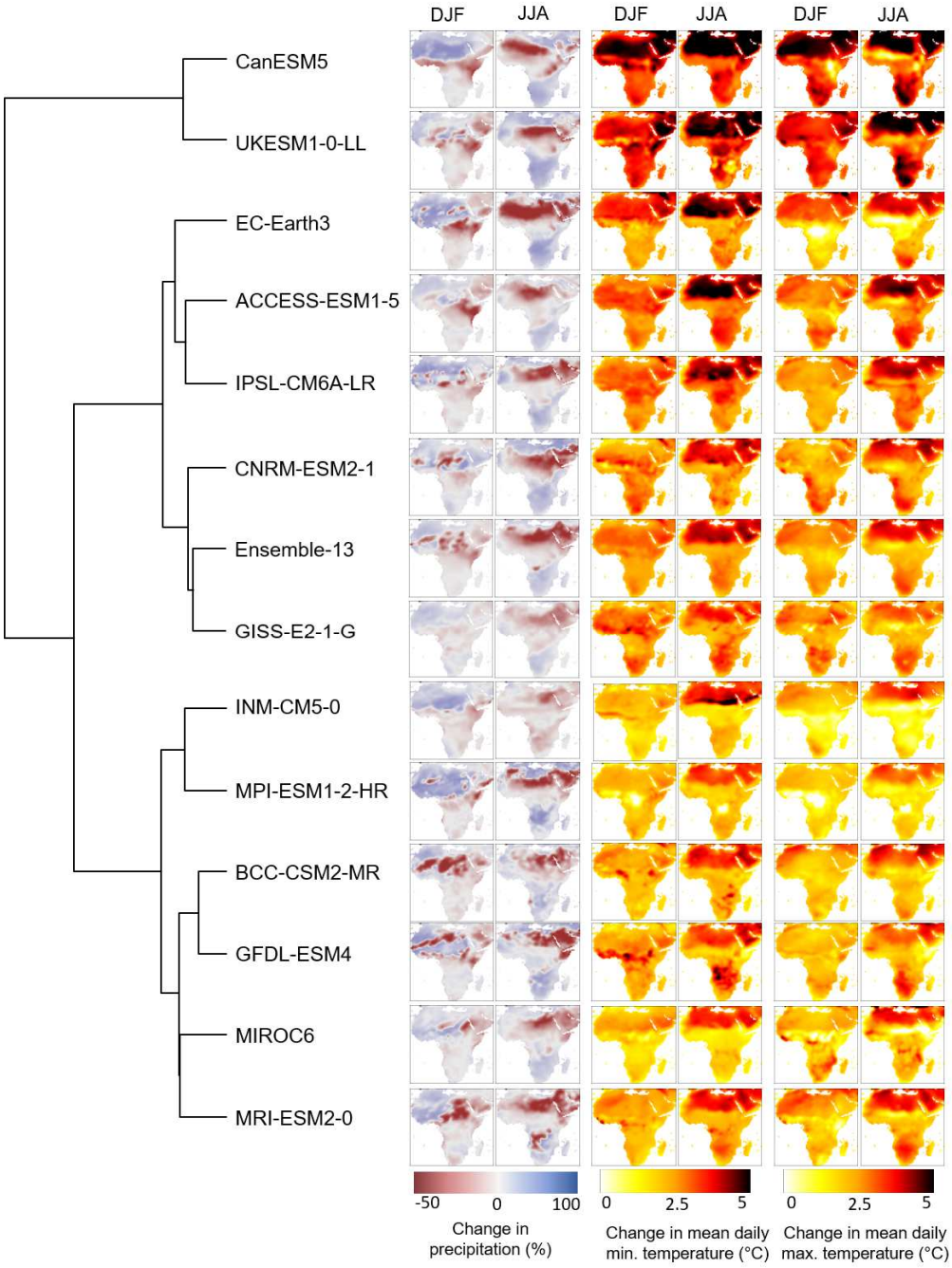


Figure 6: Models are structured by a cluster dendrogram showing spatial similarity in the projected seasonal changes for Tmin, Tmax and precipitation in December, January and February (DJF) and and March, April, May (MAM) in the period 2041-2070 under (SSP2-4.5). The maps illustrate the visual changes across the African continent for this period. Precipitation is log-scaled to provide proportional magnitude of positive and negative changes. [Problem with apparent “cooling” in bottom 4 models needs fixing]

Table 3: 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. [15]. The IPCC reference regions for Africa as shown in Figure S1 include: 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; WCA, West Central Asia and; WSAF, West-Southern-Africa. For a map of the region delineations, refer to Electronic Supplement A, Fig S1, and for an equivalent table that excludes the sensitive UKESM1-0-LL scenario, see Table 5 of the corresponding paper.

Subset size	IPCC reference region											
	ARP	CAF	ESAF	MDG	MED	NEAF	SAH	SEAF	WAF	WCA	WSAF	Africa
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