Technical guidelines for using CP4-Africa simulation data

FUTURE CLIMATE FOR AFRICA

November 2020

CONTENTS

Acknowledgements....3
About the guide....4
What is a Convection-permitting model?....5
What CP4-Africa simulations have been performed?....7
How should we expect the results from CP4-Africa to differ from CMIP models?....10
What are the limitations of its use?....12
How can I get hold of data from the model?....15
Steps on accessing the data....18
Frequently Asked Questions....19

ACKNOWLEDGEMENTS

Authors

Catherine Senior, Met Office
Declan Finney, University of Leeds
Zablone Owiti, SouthSouthNorth
Dave Rowell, Met Office
John Marsham, University of Leeds
Lawrence Jackson, University of Leeds
Segolene Berthou, Met Office
Elizabeth Kendon, Met Office
Herbert Misiani, IGAD Climate Predictions and Applications Centre

Suggested citation

Senior, C., Finney, D., Owiti, Z., Rowell, D., Marsham, J., Jackson, L., Berthou, S., Kendon, E., Misiani, H. (2020). Technical guidelines for using CP4-Africa simulations data. Future Climate for Africa. https://doi.org/10.5281/zenodo.4316466.

Cover photo: Jeremy Bishop - Unsplash

ABOUT THE GUIDE

This guide provides a practical overview of the first pan-African, kilometre-scale convection-permitting regional climate simulations (CP4-Africa), run as part of the Future Climate for Africa (FCFA) programme's Improving Model Processes for African Climate (IMPALA) project. CP4-Africa provides the first convection-permitting resolution, multi-year climate simulations for present-day and idealised future climates on an African-wide domain. The simulations have provided an unprecedented level of climate detail across Africa and initial studies have shown improvements in the simulation of many, but not all, aspects of African climate.

The goal of this guide is to promote adoption of the CP4-Africa approach within the climate community and is targeted at researchers with interest in progressing this relatively new modelling approach to improve understanding and representation of the drivers of African climate. It is also targeted at potential users of the high-resolution simulations for impact studies and decision support. While this is not an exhaustive review of the CP4-Africa simulations, from a practical perspective it highlights what one has to be aware of when designing similar simulations or using CP4-Africa simulation outputs. The guide is envisaged to be a living document that will be updated as new experiences become available from further analysis of CP4-Africa simulations data; and running CP4-Africa simulations under different experimental set ups.

The guide is structured as a series of questions on CP4-Africa simulations and the application of the results, including how to access the simulations' data. It begins with background information on what convection-permitting models are. A description of the CP4-Africa simulations performed by the IMPALA project is then presented. This is followed by an illustration of what is new in the CP4-Africa regional climate simulations, its limitations and how the simulations' data can be accessed. Finally, case studies on user experiences in accessing and using CP4-Africa simulation are presented.

About FCFA

Future Climate For Africa (FCFA) is a £20 million programme funded by the UK Foreign, Commonwealth and Development Office (FCDO) and Natural Environment Research Council (NERC). It is generating fundamentally new climate science focused on Africa and piloting the use of improved medium- to long-term (5 – 40 year) climate change information in development projects. FCFA is made up of five international research consortia and a Coordination, Capacity Development and Knowledge Exchange (CCKE) unit. Research was carried out by the following consortia:

AMMA-2050 (African Monsoon Multidisciplinary Analysis 2050)

FRACTAL (Future Resilience for African Cities and Lands)

IMPALA (Improving Model Processes for African Climate)

HyCRISTAL (Integrating Hydro-Climate Science into Policy Decisions for Climate-Resilient Infrastructure and Livelihoods in East Africa)

UMFULA (Uncertainty Reduction in Models for Understanding Development Applications)

WHAT IS A CONVECTION-PERMITTING MODEL?

Climate models are primary tools used to estimate how climate might change in future (read more on "How to understand and interpret global climate model outputs"). They are able to simulate many of the most important atmospheric and oceanic physical processes and the internal feedbacks within the climate system, however, some persistent challenges still remain. Rainfall, one of the most important variables to assess in a changing climate, is not adequately simulated in present climate models. For Africa, the ranges of available climate models largely disagree in the direction of projected changes (i.e. whether getting wetter or drier). Reliability of model estimates also reduces from larger to smaller scales as their ability to accurately represent local climate influences decreases. This is partly due to the inability of models to explicitly simulate key small-scale processes like thunderstorms and the effects of orography such as mountain ranges on local climate variations.

Global and regional climate models cannot simulate local-scale processes such as convection due to their relatively large grid cells (coarse resolution), typically several tens of kilometre across. They can simulate large-scale climate processes that operate at low or coarse spatial resolutions (>>10 km²). These models represent the average effects of convection through some form of parameterisation[1]. This simplification is a known source of model error, especially in the tropics where convection is integral to circulation and extremes.

The inability of global and regional models to accurately capture local-level processes results in biases in key outcomes such as temperature and rainfall compared to observations that increase uncertainty in future climate projections from these models. Further, many of the most important impacts of climate change on society can typically be found on the micro- and meso-scale. For example, water supply management demand for reliable climate projections on the scale of single river catchments are in most cases much smaller than the resolution of modern Global Climate Models (GCMs). All processes which have smaller spatial scales than those resolved in the GCMs cannot be represented explicitly.

[1] A method of representing processes that are too small-scale or complex to be physically represented in the climate model by simplified equations. Examples of such processes include the descent rate of raindrops, convective clouds, simplifications of the atmospheric radiative transfer and cloud microphysics

As improvements in technology and data sharing allow researchers to set up novel experiments to overcome the problem of large grid cells, scientists have developed very high-resolution climate models with grid cells that are a few kilometres wide, rather than tens of kilometres. They are known as "convection-permitting" models because they can simulate larger convective storms without the need of parameterisation schemes. They have been shown to improve the representation of dynamics such as the influence of mountains and statistical properties of convection and heavy rainfall, and hence have the potential to better represent changes in convection and local storms in future projections.

The differences between traditional GCMs, Regional Circulation Models (RCMs) and convection-permitting (CP) models are provided in the table below. CP models enhance understanding of local high-impact weather (HIW) events. The potential added value of convection-permitting model is greatest (i) at short time scales, (ii) when convection is the dominant cause of rainfall, and (iii) in regions of complex landforms (e.g. mountains, valleys and lakes).

Table 1: Differences in model properties

Model/Properties	GCM	RCM	CP
Grid-spacing	Low (>100km)	Medium (10-50km)	High (1-4km)
Convection representation	Based on simplified formulas	Based on simplified formulas	Resolved explicity
Representation of local landforms (e.g. mountains and valleys) that influence local climate (see Figure 1)	Poor	Resolve some landforms	Resolve most landforms
Representation of timing, duration and spatial distribution of sub-daily rainfall	Poor	Some improvement	Significant improvement
Representation of localised short-duration rainfall extremes	Poor	Some improvement due to orographic influence	Significant improvement
Resolve convection plumes, small showers and shallow clouds	No	No	No

Since convection-permitting models are better at characterising local-scale rainfall generating processes, they promise improvements in estimating local impacts due to HIW events. Their capacity to represent convection provides an opportunity to study many local-scale processes that have profound impacts on the economy and citizens' livelihoods. These include critical factors that influence water cycles, such as convection systems in the Congo Basin in Central Africa, where the world's most intense thunderstorms occur, the Lake Victoria circulation system in East Africa, and the West African monsoon region where convection is most organised.

Convection-permitting simulations are already being used in other regions of the world, such as in the UK, for operational forecasting and climate research. However, this is not yet the case in Africa, except for the South African Weather Service that runs convective-scale simulations to assist with forecast operations across the southern part of Africa. Under the FCFA programme, the IMPALA project conducted the first simulations of a convection-permitting model for the whole African continent (CP4-Africa). The cloud-system-resolving simulations, based on the Met Office Unified Model (UM), were designed to improve understanding and representation of the local processes that modulate African climate and hence reduce uncertainty in future projections.

The analysis of the 10-year pan-African, high-resolution (4.5 km) convection-permitting simulations under the FCFA programme has demonstrated how important simulation of convective processes are for modelling African climate today and in the future. However, the simulations have only been performed with one driving model, a single set of experimental parameters, idealised soils, fixed aerosols and consideration of only one Representative Concentration Pathway (RCP), in this case, RCP8.5 due to computational and time limitations. There is, therefore, a need to extend this work to multiple (ensemble) experiments with different models or model versions over a longer time-period for robust confidence and sensitivity assessments; and to promote the approach to a wider African science community and encourage comparison of results.

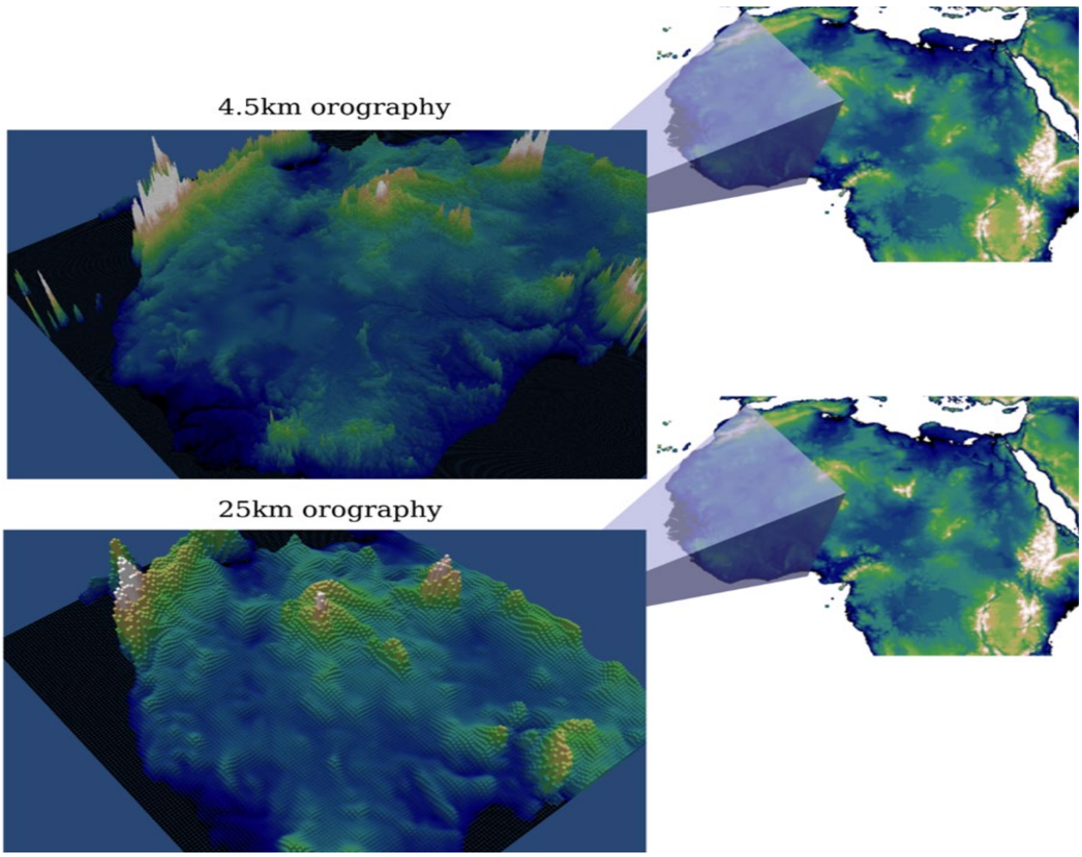


Figure 1. Illustration of improved representation of local landforms that influence local climate at convection-permitting resolution

WHAT CP4-AFRICA SIMULATIONS HAVE BEEN PERFORMED?

Two sets of atmosphere-only simulations have been performed, one for the current climate and one for the future climate. The simulations consisted of two 10-year periods, one corresponding to the present-day climate (1997–2006) and the other to an idealised representation of future climate around 2100, under RCP8.5. Each set of experiments consists of a global model run using a prototype version of the Global Atmosphere 7.0 (GA7) and Global Land 7.0 (GL7) configuration; the latest science configuration of the Met Office Unified Model (Walters et al 2018). The global model produces output fields to drive two regional models covering a domain over the African continent. Figure 2 shows the terrain heights (in metres) and the region for the CP4-Africa model.

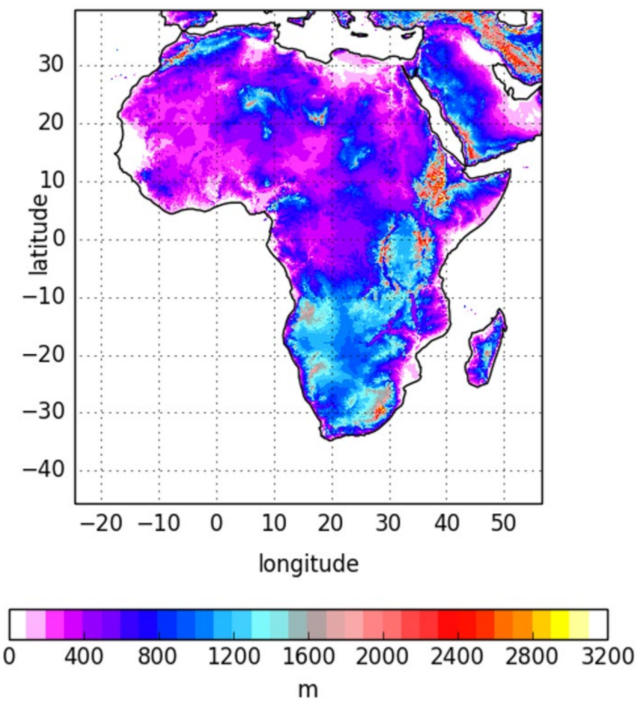


Figure 2: Domain of theCP4-Africa model showing the orography as represented by the 4.5 km grid. The R25-Africa model is run over the same domain at 25 km grid-spacing.

The first of the regional models is the convection-permitting model - CP4-Africa (Stratton et al 2018) with a 4.5 km grid spacing. The second regional model is R25-Africa, a regional model based on the global model, with the same ~26 km latitude and ~39 km longitude grid spacing across Africa as the global model, but over the same regional domain as CP4-Africa. R25-Africa was run to aid in the understanding of the differences between the CP4-Africa and global model, in particular, to isolate the impact of the convection parameterisation and the influence of the land-surface boundary condition. The major difference between the regional simulations, apart from horizontal resolution, is that CP4-Africa runs without any convection parameterisation, whereas R25-Africa uses the same convection parameterisation as the global model. Full details of the other model differences between CP4-Africa and R25-Africa are given in Stratton et al 2018.

Experiment design for current climate simulations

The global model simulation covers around 30 years from 1 September 1978 until 1 December 2010. The regional climate simulations run for just over ten years from 1 January 1997 to 1 March 2007, are forced by global model fields and so inherit biases from the global model. The atmospheric initial conditions for both the CP4-Africa and R25-Africa simulations are taken from global atmospheric model fields for 1 January 1997. The CP4-Africa and R25-Africa models are forced by one-way nesting with lateral boundary conditions derived from the global atmospheric simulation as shown in the schematic below. Due to some problems early in the simulation and the cost of the model, the CP4-Africa simulation was run in various sections as shown in the schematic in Figure 3 below. Thus, the current and future climate simulations have multiple run IDs. The schematic diagram lists the run IDs (i.e. u-plus a five alpha-numerical id) used for different parts of the simulation, while the table provides suites IDs required for retrieving the data from Managed Archive Storage System (MASS). The last part labeled as u-aj575 in the diagram was initialised from CP4-Africa model conditions from 1 January 2000 relabeled as 1 January 2004 from run u-ah261. The future data labelled 1997-2006 represent a decade around the year 2100 (2097 - 2106).

The soil properties for both CP4-Africa and R25-Africa were defined to be spatially uniform (and those of sand) across the whole domain; with the soil moisture fields initialised with climatological data derived from an off-line JULES land surface simulation on a 0.5-degree grid. This was forced with a bias-corrected reanalysis data set: the WATCH Forcing Data 2013 ERA-Interim (Weedon et al 2014), in which the monthly air temperature and rainfall totals are bias corrected against GPCC and CRU TS3.1 gridded observations. Initialising the soil moisture in this way is relatively fast and has the advantage of ensuring the soil moisture in all four soil layers is adequately spun up.

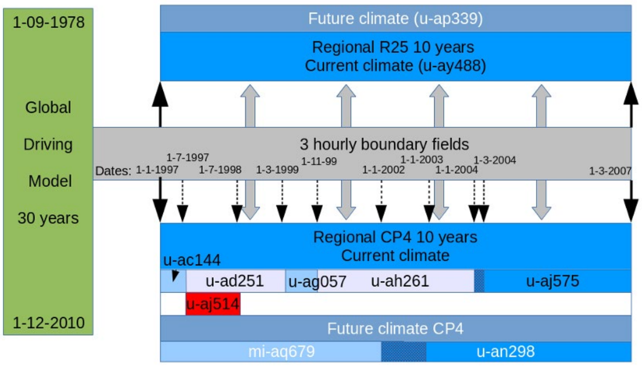


Figure 3. Schematic diagram of control climate simulations

Table 2: CP4-Africa simulations suites IDs

Simulation	Suit ID	Description
25 km regional control	u-ay488	10 years from 01/01/1997 to 01/03/2007
25 km regional future time slice	u-ay619 (replaced u-ap339)	10 years from 01/01/1997 to 01/03/2007
CP4 regional control	u-aj514	01/07/1997 to 01/07/1998 (corrected rerun, no flood, 75s time step)
CP4 regional control	u-ac144	01/01/1997 to 01/07/1997 (100s time step)
CP4 regional control	u-ad251	01/07/1997 to 01/03/1999 (flood at the start of this, 100s time step)
CP4 regional control	u-ag057	01/03/1999 to 01/11/1999 (100s time step)
CP4 regional control	u-ah261	01/11/1999 to 01/03/2004 stop date (75s time step)
CP4 regional control	u-aj575	01/01/2004 to 01/03/2007 (75s time step)
CP4 regional future timeslice	mi-aq679	01/01/1997 to 01/03/2002
CP4 regional future timeslice	u-an298	01/01/2002 to 01/03/2007

Surface properties and forcing

- Land-sea masks were created from the IGBP (International Geosphere and Biosphere Programme) land classification data set.
- The surface orography for the models is created from the GLOBE (Global Land One-km Base Elevation) data set. A minor modification was applied to the surface orography used for CP4-Africa in the region around Mount Cameroon after six months of simulation. At the same time as this modification, the soil moisture was unfortunately reset to a saturated value in u-ad251. Run u-aj514 correctly continues from u-ac144 without a soil moisture reset.
- Soil properties for CP4-Africa and R25-Africa were both set to be spatially uniform, with the characteristics of sand, across all land in the domain. This is different from the global driving model which has varying soil properties.
- Land cover fractions are derived from the European Space Agency climate change initiative: the land cover data set (CCI-LC) version 1.3 (Poulter 2015) for the epoch 1998 to 2002.
- The leaf area index is updated every five days using a monthly climatology created from MODIS Collection 5 mapped to the five plant types used in the land surface 9-tile scheme.
- All the models are forced with SSTs derived from the Reynolds data set of daily high-resolution blended analyses for SST (Reynolds 2007). These data have a spatial grid resolution of 0.25 degrees and are interpolated onto the regional model grids using bi-linear interpolation.
- In both CP4-Africa and R25-Africa, land-sea mask has numerous lakes which are represented as inland sea points, the majority being in East Africa, with the largest being Lake Victoria, covering 3502 grid boxes on the CP4-Africa grid. Where lakes are included in the ARC-Lake v3 data set (Hook 2012, http://www.geos.ed.ac.uk/arclake/documents.html), a climatology from this data set of monthly night-time lake temperatures has been used. For other lakes, typically those with a surface area of less than 50 km² a surface temperature value from the model's nearest sea point is assumed.

The radiation and cloud microphysics schemes require three-dimensional fields of ozone mixing ratio, aerosols and dust particles. Climatological values (see Stratton et al 2018 for full details) are assumed for all, and these are updated in the model every 5 days. Various greenhouse gases are assumed to have fixed global values, which are varied annually over the 10-year simulation. Carbon dioxide mass mixing ratios are varied from 0.5551679 g/kg for 1997 to 0.581488 g/kg for 2006 in the same way as in the global model.

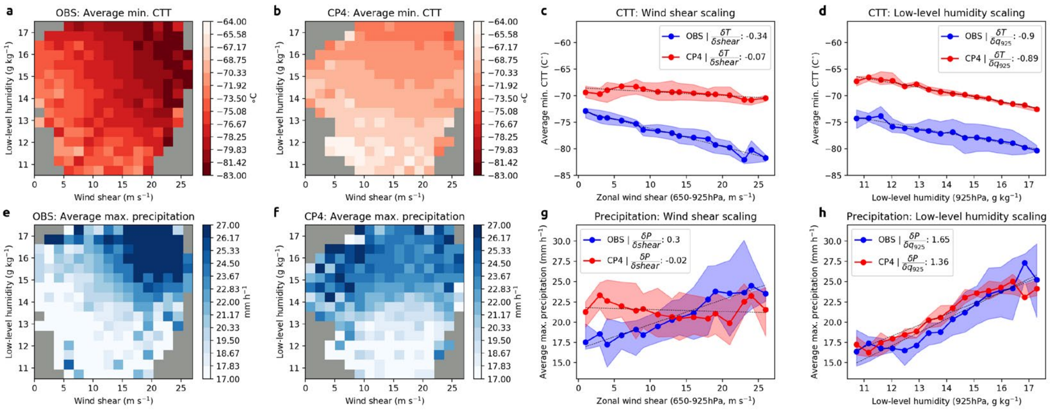


Figure 4. Example of CP4-Africa model capabilities in simulating MCS intensity scaling with atmospheric drivers from Senior et al.2020. Histograms shows average minimum cloud top temperature (CTT) of (a) observations (OBS) and (b) CP4-Africa, and average maximum precipitation of (e) OBS and (f) CP4-Africa as a function of pre-storm specific humidity and absolute zonal wind shear at storm location for afternoon MCSs over the Sahel. Line plots depict the average (c,d) minimum CTT and (g,h) maximum precipitation for OBS (blue) and CP4-Africa (red) associated with each bin of (c,g) wind shear, where shading spans the inter-quartile range (IQR) of humidity within the shear bin, and (d,h) humidity, where shading spans the IQR of shear within the humidity bin.

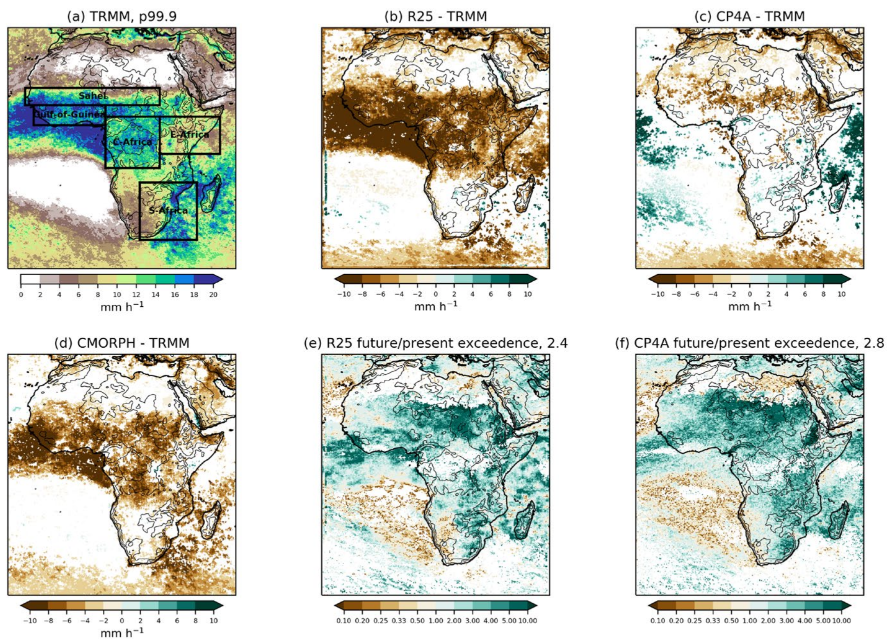
Experiment design for future climate simulations

The future simulations correspond to a 10-year period around 2100, for the IPCC RCP8.5 climate change scenario.

Future forcings are specified following a similar approach to that of the UPSCALE project. Namely, the SSTs are the sum of the SSTs used in the present-day simulations and the climatological average SST change between 1975–2005 and 2085–2115 in a HadGEM2-ES RCP8.5 run. These SST changes were calculated for each calendar month, interpolated in both space and time, and added to the daily varying Reynolds forcing data on the various model grids. The increase in SST forcing equates to a global mean SST increase of just under 4 K, giving a global mean 1.5 m air temperature change of 5.2 K for the period of the CP4-Africa simulations. The same ozone and aerosol climatologies are used in both the future and present-day simulations. Greenhouse gas values were taken from the RCP8.5 climate change scenario for the year 2100.

Lake surface temperatures for the future CP4-Africa simulation were computed as the sum of the ARC-Lake observations and a seasonally varying change in Lake Surface Temperature (LST) specified from the N512 GCM 1997–2007 and the corresponding time period in the future simulation monthly climatologies. All African lakes in N512 GCM, except for Lake Victoria, are land points. For more details on future climate simulation design see Kendon et. al. 2019.

Figure 5: Example of CP4-Africa model capabilities in representing present-day extreme precipitation and the frequency of exceedance in future from Kendon et al. 2019. Present-day extreme precipitation threshold in (a) TRMM observations, differences with respect to TRMM for (b) R25 model, (c) CP4A model, (d) CMORPH observations, and the ratio of the future compared to the present-day frequency of exceedance of this threshold for (e) R25 model and (f) CP4A model. Extreme precipitation threshold is defined as the 99.9th percentile of 3-hourly precipitation in the wet season in the present-day. The median of future/present exceedance ratio across Africa is indicated in (e) and (f) panel titles.



HOW SHOULD WE EXPECT THE RESULTS FROM CP4-AFRICA TO DIFFER FROM CMIP MODELS?

The 4.5 km resolution and the explicit representation of convection in CP4-Africa means we expect to see improvements in the representation of both regional-scale circulations and small-scale climate processes compared to coarser-resolution simulations with parameterised convection. A regional simulation with parameterised convection (R25-Africa) was used to interpret the differences between the convection-permitting simulation (CP4-Africa) and the global driving model. The R25-Africa simulation has been used to compare the explicit convection simulation (CP4-Africa) to a very similar simulation where convection is parameterised (R25-Africa). Many aspects of the simulation have been kept constant in both cases such as boundary conditions and GHG concentrations. However, as well as the difference in representation of convection, R25-Africa also runs at a coarser resolution and has a different cloud microphysics scheme.

The parameterised convection approach is typical of the method used in CMIP and CORDEX models, though applied at a relatively high resolution (~25 km). The R25-Africa model uses the same physics as the global driving model but provides a better comparison to the CP4-Africa simulation as it uses the same domain, land surface, and aerosol climatologies. Some of the major differences between CP4-Africa and typical CMIP models are related to the changes in atmospheric moisture and stability and are described below. In many cases, these results confirm earlier findings with convection-permitting models over smaller regions of Africa (e.g. found under the Cascade project; Marsham et al 2013):

1. Widespread improvement in the rainfall intensitiesand hourly rainfall characteristics in CP4-Africa simulations (Stratton et al 2018; Kendon et al 2019). Increases in sub-daily precipitation extremes in CP4-Africa are due to higher scaling rates of atmospheric moisture in convection-permitting models. In West Africa, for example, the diurnal cycle of rainfall is improved due to better representation of the life cycles and direction of propagation of meso-scale convective systems (Berthou et al 2019, Crook et al 2019). However, in the Sahelian zone away from the orography, small-scale afternoon/evening convection contributes too much total precipitation. Therefore, the afternoon peak is prominent in all regions whereas in the satellite observations some regions have early morning peaks (ibid). The storms in the CP4-Africa simulations have more realistic diurnal cycles, lifetimes and propagate in the correct direction and have much improved spatial distribution, but still have too few large systems (Crook et al 2019). In East Africa, simulation of the local terrain and convective processes improves the daily rainfall cycle and severe storms over the basin. These include a more accurate representation of the proportion of dry 3-h periods; reduced bias in the representation of the diurnal cycle in parts of the Kenyan and Ethiopian highlands; and a shift in the mean rainfall pattern from Congo to the Lake Victoria basin due to differences in large scale moisture fluxes - resulting from continental-scale impacts of explicit representation of convection through modification of sources, sinks and transport of moisture over East Africa (Finney et al 2019).
2. CP4-Africa simulations also show stronger meridional overturning into subtropical southern African compared to simulations run with lower resolution parameterised convection models (Hart et al 2018). This leads to an improved representation of the annual cycle of tropical-extratropical (TE) cloud bands, which are fundamental to the regional hydroclimate and extremes there. Under climate change, stronger interactions between the convective and larger-scale flows are apparent in CP4-Africa compared to the parameterised convection model (Jackson et al 2020). In the tropical rain-belt, CP4-Africa has stronger couplings between its changes in distributions of vertical velocity, rain intensity and TCW. These couplings contribute to a greater slowdown in mean Hadley ascent and a weaker increase in mean rainfall along with a greater shift to more intense convective-scale updraughts and rain intensity.
3. Changes in extreme rainfall and dry spells over Africa may be underestimated in traditional models where convection is parameterised. CP4-Africa simulations suggest that heavy rainfall events (exceeding 60mm rainfall in 3h over 25 km x 25 km area) will increase at a greater rate than in models

that parameterise convection (e.g. in the CMIP and CORDEX ensembles) and extreme rainfall events that occur approximately once every 30 years now, may be once every 3-4 years by the end of the century, under RCP8.5 (Kendon et al 2019; Berthou et al 2019; Finney et.al 2020). Dry spells during the wet season exceeding 10 days in length are almost twice as frequent in the future compared to the present-day: a signal which is again not seen in a coarser resolution parameterised model. The future changes in extremes over Africa may be more severe than previously thought.

4. Representation of lightning in CMIP models is highly simplified, if present at all. The R25-Africa model also does not include a representation of lightning, but the CP4-Africa model includes a lightning scheme much more closely linked to physical processes than schemes typically used in CMIP models (Finney et al 2016). The lightning simulated by the CP4-Africa model has been shown to compare well to observations, and its projections have been described and interrogated by Finney et al (2020b).

5. The different cloud microphysics schemes do not prevent seeing clear differences in the models regarding the different representations of convection. However, it is important to bear in mind that it will affect the simulation of climate. For instance, the CP4-Africa cloud scheme generally produces more optically thick clouds, which leads to a cooler surface air temperature. Around Lake Victoria this leads to a different lake-land contrast and therefore convergence onto the lake (Finney et al 2019).

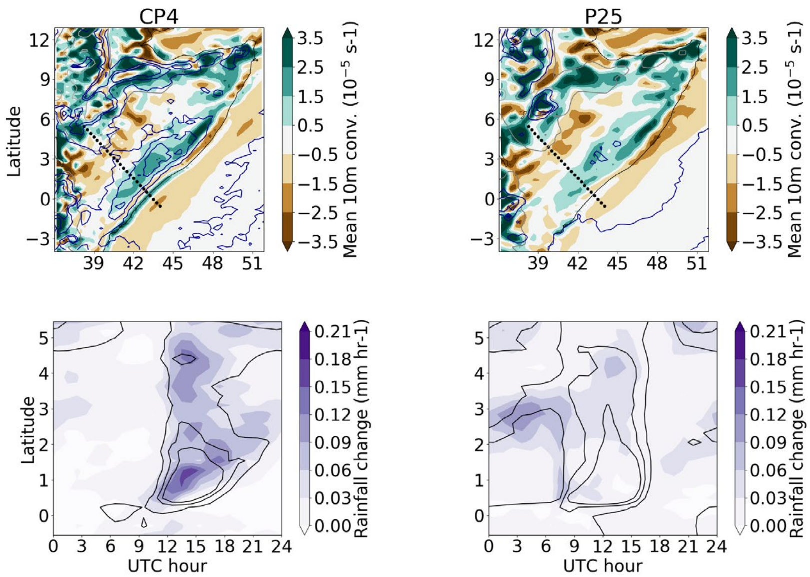


Figure 6: Example of CP4-Africa model capabilities in simulating sea breezes from Finney et al. 2020. Transect Hövmöller plots of rainfall and dynamical changes over the Horn of Africa for (left) CP4 and (right) P25: (top) mean 1800 EAT 10-m wind convergence and rainfall (blue contours), along with the black dotted transect used in the lower panels; (bottom) mean changes under future climate, where solid contours show the current climate (rainfall propagation in current current is improved in CP4A vs P25, not shown).

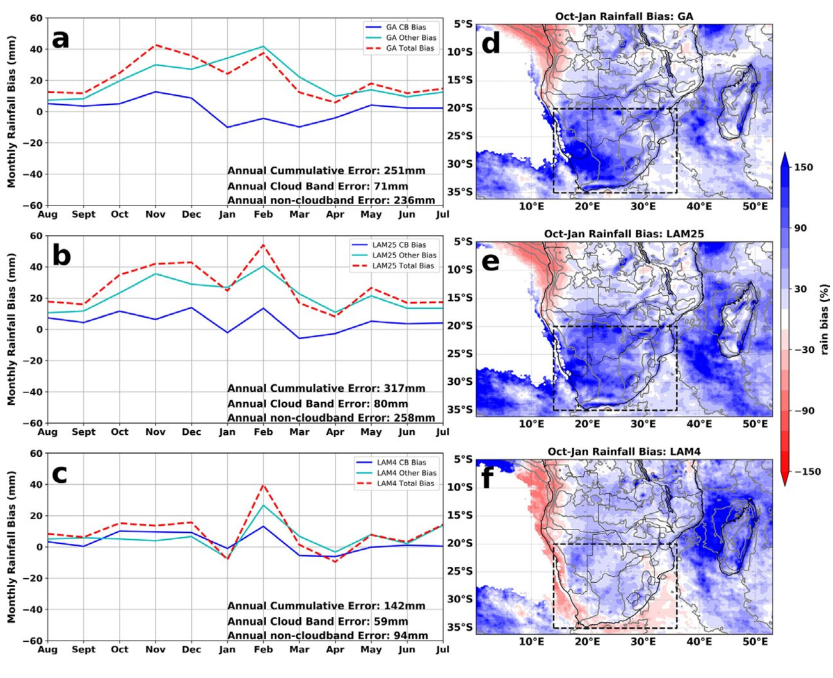


Figure 7: Example of CP4-Africa model capabilities in simulating cloud band rainfall from Hart et al. 2018. Monthly area-averaged (14°E-36°E, 20°E-35°S) rainfall bias with respect to TRMM rainfall observations simulated by (a) GA N512, (b) regional LAM25, and (c) regional convective-permitting LAM4 (CP4-Africa). Total bias (red dashed) is decomposed into bias due to TE cloud band rainfall (dark blue) and bias due to rainfall from other systems (cyan). The accumulated annual error (sum of monthly absolute error) is given in the panel text for each of the simulations. Maps of total bias (% of climatological rainfall) in October to January rainfall simulated by (d) GA N512, (e) regional LAM25, and (f) regional convective-permitting LAM4. Domain for area averages in a-c indicated in black dashed box. Areas with annual rainfall <10 mm are masked. GA = Global Atmosphere; LAM = limited area model; TRMM = Tropical Rainfall Measuring Mission; TE = tropical-extratropical.

WHAT ARE THE LIMITATIONS OF ITS USE?

While CP4-Africa has enabled practical investigation of the importance of representing local, small-scale weather features explicitly in the simulation of climate and projections of climate change across any part of the continent, it has a number of noteworthy limitations from its experimental set up and other aspects that need to be taken into account when using the data and/or interpreting the results.

1. Experimental setup

- *One Representative Concentration Pathway* (RCP) – in this case, RCP8.5, which leads to around 4-7 degrees warming of surface air in the future climate. This experiment is limited in its potential to provide information regarding the uncertainty of climate change with respect to different concentration pathways that humanity may take. The choice of RCP8.5 which represents the large change in climate we are heading for now does, however, provide a strong forcing giving a better chance than other RCPs of observing a clear signal.

- *One time slice comparison* – In this case, years representative of 2097-2106 are compared to 1997-2006. Therefore, the simulations do not provide direct information about other time periods. Like the choice of a high-end RCP, the choice of the end of the century is to deliver a clearer signal of climate change. To deliver mid-century projection information to users, two approaches have been taken in FCFA. One has developed a statistical methodology that scales the distribution of daily CP4-Africa data by the HadGEM2-ES evolution of global annual mean temperature, combining this with CMIP and observational data for a 'site-specific synthesis of the projected range' (Mittal et al 2020, submitted). The other approach develops physically-based regional relationships between large-scale climate drivers and storm characteristics – derived from CP4-Africa and observational data – to scale CMIP data to projections of the evolution of regional storm intensity throughout the 21st century. We note that SST patterns are shared between present and future, however, this leads to much shared interannual variability between the two periods, in at least some regions (Wainwright et al, 2020)

- *A regional model* – The convection-permitting simulation is applied to a limited area encompassing Africa. A global model, using a coarser resolution and convection parameterisation is used to provide boundary conditions to the regional model. These are not two-way coupled, i.e. the regional model does not feed back to the global model, and the regional model is only free to evolve away from the parent global model's boundary conditions, which act to constrain it.

- *One driving model* – In this case the UK Met Office Unified Atmosphere-only Model (GA7), N512 resolution, including convection parameterisation. Due to computational restraints, only one driving model is used. Therefore, this dataset does not provide direct information regarding the uncertainty of the simulation of the global climate, and its influence on African climate. However, it provides indirect information, for example, on how convection can affect large-scale climate change.

- *One ensemble member* – Variations in initial/boundary conditions or model parameters can provide a range of results which helps to infer the internal variability of the simulated climate, and can therefore improve the statistical robustness of the analysis. Due to computational restraints, only one realisation has been simulated and therefore it is more difficult to determine whether results are significant. However, the use of a high greenhouse gas scenario and long time period should ensure signals are clearer.

- *10 years of simulation* – This time period allows for a reasonable quantification of the mean climatology and effects of climate change. Furthermore, some analysis of extreme events can be performed but these are currently limited to approximately 1-in-10-year events, or for statistical significance, likely more frequent events. Careful consideration of sample size should be given if applying analysis other than long-time averages.

- *No aerosol change* – Only sea-surface temperature and well-mixed greenhouse gas concentrations are altered for the future climate. This allows for clear attribution of the most fundamental long-term climate change but does not provide information regarding changing aerosols which, in reality, will modify climate change

- *Prescribed sea and lake surface temperatures* – This model simulates the atmosphere and the land. However, water bodies are boundaries which have prescribed temperatures that vary monthly. This allows attribution of changes to atmosphere and land processes, but does not provide information regarding the interaction of the atmosphere and the ocean or lakes, which may dampen or enhance climate changes. See Finney et al (2019) for more detail on the role of Lake Victoria in the simulation, and Finney et al (2020) for results and discussion regarding the prescribed lake surface temperature changes in the future climate simulation. In particular, prescribed SSTs may be linked to the overprediction of intense rainfall over oceans.
- *Sandy soil* – All soil in the land component of the model has the same sandy soil properties. This reduces biases that can result in the atmospheric coupling as a result of inaccuracies in the soil properties ancillaries. However, it may mean that in some locations surface moisture and radiation, and soil properties, may not be entirely representative.

2. Resolution

CP4-Africa runs at a very high resolution relative to typical global and regional climate models. It does so to explicitly simulate convection. However, convection happens on many scales, and the smaller scales (<4.5 km) will not be represented here. As such, there will likely not be a good representation of small cumulus clouds and, therefore, not a full representation of the formation of cumulus to deep cumulonimbus. This may impact the radiation and moisture balances at the initiations of simulated storms. However, work to date on the CP4-Africa dataset has shown the resolution to be appropriate for drawing conclusions regarding well-formed storms and extreme precipitation rates, for example meso-scale convection system size and frequency is well represented in the Sahel.

Nevertheless, if CP4-Africa is analysed on its native 4.5 km grid, precipitation is generally too intense (low precipitation rates are underestimated and high rates overestimated), for example during July-August in the Sahel (Berthou et al 2019). This is a common problem amongst convection-permitting models (Pichelli et al 2020 under review) because deep convection is not perfectly represented at the 4 km scale and turbulence parameterisations are in a grey-zone (Hanley et al 2015) and meso-scale convective systems tend to be too small and intense (Crook et al 2019). Therefore it is not recommended to use CP4-Africa precipitation data at its native grid (or that a bias correction is applied to its intensity distribution). If instead CP4-Africa is aggregated to a 25 km grid, the precipitation distribution compares well to observations over the northern Sahel, for example. However, low to moderate precipitation rates are still underestimated and heavy rainfall is too intense in regions where processes at smaller scales than the meso-scale convection systems play a greater role (eg. the Guinea Coast and Soudanian zones), and bias correction is likely still needed for many applications.

That said, CP4-Africa may still contain usable and robust sub-25 km information in regions of large surface variability, e.g. around large mountains, narrow ridges, lake coastlines or urban tiling. Further work is assessing the location and circumstances of any such regions. Furthermore, CP4-Africa's high resolution in mountainous regions, eg. East Africa and around Mount Cameroon, allows for a much better representation of surface elevation. Even so, the steepest gradients and highest peaks may still differ notably from reality, so that comparisons with point observational data should be considered carefully.

3. Incorrect latitude/longitude values

The model was run with orography and a land-sea mask that is offset by half a grid-length from reality. The latitude and longitude metadata of the orography and land-sea mask ancillary files were not those used by the model. But the metadata in other CP4-Africa output files do contain the latitudes and longitudes used by the model. This issue must be addressed in any analysis that requires accurate referencing of atmospheric fields (or land-surface fields computed during the model simulation) against topographic or land-sea locations. So in the reference frame of the atmospheric data, the latitude and longitude metadata of the orography and land-sea mask ancillary files should be adjusted by half a grid-length to match that of the atmospheric data. Note also that in this reference frame, this issue infers a very small mismatch between the topographic or land-sea locations used by the model and the model dynamics (Coriolis force) or diurnal cycle (solar time).

4. Simulation continuity

For various reasons, there are three times in the CP4-Africa simulations at which the simulation does not run on from the previous output files (Figure 3). These are:

Current climate CP4-Africa:

- 1st July 1998
- 1st March 2004

Future climate CP4:

- 1st March 2102

This will have little effect on the majority of analysis. It is relevant though for analysis regarding storm tracking; storms on the dates listed above will not follow on from the previous day. This again should have little impact on storm-tracking analysis if properly considered in the tracking.

5. Data analysis difficulties and challenges

There are a few peculiarities with the data that are worth noting:

- *Missing file*: there is a missing file for all diagnostics for one hour on 21/6/2003 current climate CP4-Africa. This file got corrupted during production. Other hours on that day are available.
- *15-minute precipitation data exist* (b04203): but they should not be used for data preceding March 1998 in the future climate CP4-Africa simulation. The diagnostic was not correctly set up for the first 14 months.
- *Pressure-level data*: There is pressure-level data missing early on in some runs until September 1997. There is also a switch where 60hPa level exists early on to 600hPa later. The 600 hPa level is available for the following simulation dates: 01/Jul/1997-30/Jun/1998 and 01/Jul/2000-30/Feb/2007.
- *CP4 current and future climates have multiple run IDs*: the details of this have been discussed, but be aware that this can cause clashes in metadata if using Python/Iris.

HOW CAN I GET HOLD OF DATA FROM THE MODEL?

CP4-Africa output includes two main datasets: a 10-year simulation of present-day climate 1997-2006; and a 10-year idealised simulation of future climate representative of 2100. The CP4-Africa and R25-Africa datasets generated under the FCFA IMPALA project are available via:

1. The Centre for Environmental Data Analysis (CEDA) archive here. A guide on how to register and access the data is provided here. This publicly available archive includes a limited set of the most widely-used variables (monthly mean and hourly data). To access the CP4-Africa and R25-Africa data (labelled P25 on the CEDA archive) please type in 'CP4-Africa' in the search box on the CEDA home page. You will not need to register to download these datasets.
2. The Joint Analysis System Meeting Infrastructure Needs (JASMIN). Documentation for all aspects of the JASMIN scientific data access and analysis environment can be accessed here. The full datasets for CP4-Africa and R25-Africa are stored on the Met Office MASS archive system. This can be accessed via JASMIN but you will need to seek permission for access to the relevant user IDs. Please read the case studies below for a detailed account of how to go about using the complete datasets.

Case study of user experiences in accessing and using CP4-Africa data

a) *UK ACADEMIC COMMUNITY: DECLAN FINNEY, HyCRISTAL PROJECT*

Introduction

Part of my role as HyCRISTAL postdoc at the University of Leeds was to access CP4-Africa data through the JASMIN and MASS computer systems, and analyse this data over the East Africa region. I also helped facilitate access for other researchers at Leeds and within the HyCRISTAL project. Below are some insights I hope will be useful for UK researchers looking to use CP4-Africa. I roughly provide the process from download to analysis, with details left to other parts of the guidance documentation. The purpose is to give an easily readable starting point from which researchers can look into the details as and when they come to need them.

Accessing data

There are three ways a UK researcher might access the data, in order of easiness:

1. Existing downloaded data in institutional data storage
2. A selection of some widely-used fields from the CEDA archive
3. Downloading from the raw dataset held on the MASS archive

The first two should be straight-forward, following the guidance documentation. The MASS archive extraction is more involved and is the method I have used in my work. Anyone carrying out more detailed atmospheric or land surface analysis is likely to want diagnostics that require extraction from MASS. There is a PDF document listing the full set of diagnostics available on MASS. This data is in PP format and is very large for the 4 km simulation. It is possible to select specific files, diagnostics and model/pressure levels in order to reduce the size of the download. My investigations have not found any way to select a smaller spatial region to download, so don't be too hopeful for that. Because such large amounts of data are being downloaded, I have found it important to check the number and size of files I expect (e.g. for files not meeting certain size <<< find *.pp -type f -size –XXc >>> where XX is the expected number of bytes). This ensures the download went smoothly. If it did not, then note that there is an option to only extract 'un-downloaded or differently-sized files' from MASS. A final comment is that I found it useful to become familiar with using the screen command in Bash since it makes it easier to have long-running jobs on JASMIN.

Convert data to NetCDF

I strongly recommend post-processing the PP files to convert to NetCDF format. It can dramatically reduce the file sizes if compression is used, as well as making the files more widely usable. Conversion can be done on JASMIN (Python is one option available for this).

Where to carry out data analysis

There are two options for where to carry out analysis:

1. Use JASMIN, possibly with Lotus for bigger job submission
2. scp or rsync data to another server

I have tried the first option a little. Lotus is useful for increasing the speed of converting large numbers of PP files to NetCDF. However, I find that for my actual analysis of CP4-Africa, JASMIN and Lotus are slow and/or cumbersome compared to performing the analysis on our systems here at Leeds. This aspect is obviously highly dependent on your institution's computing facilities, so be aware there is the option of working on JASMIN.

Opportunities

You now have some data and are ready to perform analysis. Great! Well done! You probably have lots of ideas about how you want to use CP4-Africa. There is some CP4-Africa documentation describing some ways the data can be used. There are already several publications using the data from which to gather inspiration from, including my own (Stratton et al 2018; Finney et al 2019; Kendon et al 2019; Hart et al 2018; Jackson et al 2019, Jackson et al 2020; Finney et al 2020). During my analysis I have benefitted from a wide range of diagnostics that have allowed me to calculate moisture and energy budgets, composite 3-hourly data during extreme rainfall, and look at meso-scale flows at hourly timescales. For East Africa, the high resolution of the model has proven very valuable in studying processes around mountains and lakes.

Other than these, there are just the expected challenges with using such large amounts of data. I do not have any specific advice regarding this other than reading in smaller numbers of files at a time and processing before reading in more files. And saving processed data more often than might be needed with smaller datasets.

Final remarks

The CP4-Africa dataset provides an unprecedented level of climate detail across Africa. It is challenging to use and it has its drawbacks. However, as you become familiar with it, you will start to see the new and interesting angles of research that it opens up. Good luck, I look forward to seeing the research that is still to come.

Introduction

I joined the HyCRISTAL project as a research assistant at IGAD Climate Prediction and Applications Centre (ICPAC). I was assigned to one of the FCFA gap-filling projects called "Where East Meets West": the title symbolises the intersection point where convective systems that propagate westwards are generated. My main focus in the project was to analyse the CP4-Africa simulation datasets over Ethiopia and South Sudan, given that significant research using the same datasets had focused on East Africa.

Approach

The first task was to determine an appropriate tool to effectively handle the enormous CP4-Africa simulations dataset without difficulty, given that we aimed to analyse the data at sub-daily timescales. My previous experience with programming language was in R which could not effectively handle the amount of data we had beforehand. It was inevitable to switch to Python which has ready libraries that can handle such data. Through close guidance from my mentors at Leeds University, I was able to develop skills in Python to a point where I am now able to confidentially write my own functions.

Access to the Data

The computing resources required to analyse this data, in terms of processing power and storage capacity, is enormous. I therefore needed access to a cluster to perform the analysis at the timescales needed. My colleagues at the University of Leeds facilitated my access to the university cluster which already had the CP4-Africa simulations data in the repository. I was therefore able, through their guidance, to access and perform analysis remotely. Additionally, through project funds, we procured a high capacity (5TB) hard drive which we used to backup the data I needed for analysis and to make the CP4-Africa data locally available for use at ICPAC. Taking into consideration the project timelines, it was not feasible for me to handle the raw data which would have taken quite a substantial amount of time to process. I therefore used the post-processed data which was subset to the region of interest in NetCDF format.

Challenges

One of the major challenges, and which turned out to be a blessing, was learning a new programming language (Python) in a short time. However, this played out to my advantage (and my institution, ICPAC, since I now have the analytical skills to handle large sets of data. The project has therefore left a lasting legacy in me and the entire region. The next challenge was understanding the CP4-Africa data. It cost me and the project a visit to the University of Leeds, which was an eye-opener for me to fully grasp the details. Finally, the last challenge was interrupted internet connectivity given that I was working remotely. I was not able to access the resources at the University of Leeds during brief episodes of downtime. To counter this, I downloaded some of the processed data (which were significantly less bulky) to my local machine. I also had the portable hard drive with all the datasets and scripts I needed.

Opportunities

There is an opportunity for early-career scientists to use these novel climate simulations to help in generating new insights into the region's climate and its future response to climate change. Finally, there is an opportunity for ICPAC as a regional climate institution to help build the capacity of the member states to help them generate robust climate change information. This can help cement their negotiations at global climate change forums such as the Conference of Parties (COP) meetings.

STEPS ON ACCESSING THE DATA

Step 1: Get a CEDA account

You may already have a CEDA account if you have ever accessed data from CEDA, BADC, or from the CEDA CMIP5 node. If not, go to this webpage: https://services.ceda.ac.uk/cedasite/register/info/

Click "Continue" and follow the registration procedure.

Step 2: Generate and register your SSH key

Instructions on how to do this are here: http://www.jasmin.ac.uk/workflow/ssh-keys/

Step 3: Get a JASMIN account

Application for a JASMIN account is here: http://www.jasmin.ac.uk/workflow/get-jasmin-account/

Step 4: Register your network domain

Instructions here: http://www.jasmin.ac.uk/workflow/registering-your-ip-with-ceda/

Step 5: Logging in to JASMIN

Instructions on how to login are here:

http://www.jasmin.ac.uk/workflow/logging-on-to-jasmin/

There are slightly different procedures depending on whether you are using Linux, Windows, or Mac, so please see the relevant links on the above webpage. All procedures involve:

i) Adding your SSH private key

ii) Logging in to the gateway machine (the one you want is jasmin-login1.ceda.ac.uk)

iii) Logging in to the server. For this step you have a few options. If you want to do any analysis it is best to use the "sci1" or "sci2" servers:

ssh -X [username]@jasmin-xfer1.ceda.ac.uk

ssh -X [username]@jasmin-sci1.ceda.ac.uk

ssh -X [username]@jasmin-sci2.ceda.ac.uk

iv) Once logged in, you can move to the relevant Group Workspace (GWS) – probably the IMPALA GWS: cd /group_workspaces/jasmin2/impala/

How to download/upload files

Information on how to transfer data to and from JASMIN is here:

http://www.jasmin.ac.uk/how-to-use-jasmin/data-transfer/

FREQUENTLY ASKED QUESTIONS

Why are the future data labelled 1997-2006? What years do they represent?

The future simulations represent a decade around year 2100 (2097 - 2106).

What are the longitude values?

The longitudes listed in the files are the real longitudes plus 360.

What are the different run IDs, e.g. ac144? How do I take these into consideration?

For most purposes the run ID can be ignored and the whole time series of each simulation can be considered continuous. There are a few types of analysis, such as storm tracking, which may need consideration of the run IDs. For more information, look at the model description and limitation guidance notes.

Further Reading

A. CP4-Africa Simulation References

Berthou, S., Kendon, E., Rowell, D. P., Roberts, M. J., Tucker, S. O., & Stratton, R. A. (2019). Larger future intensification of rainfall in the West African Sahel in a convection-permitting model. Geophysical Research Letters, 46, 13, 299–13, 307. https://doi.org/10.1029/2019GL083544.

Berthou, S., Rowell, D. P., Kendon, E. J., Roberts, M., Stratton, R., Crook, J., & Wilcox, C. (2019). Improved climatological precipitation characteristics over West Africa at convection-permitting scales. Climate Dynamics. https://doi.org/10.1007/s00382-019-04759-4.

Crook, J., Klein, C., Folwell, S., Taylor, C. M., Parker, D. J., Stratton, R., & Stein, T. (2019). Assessment of the representation of West African storm lifecycles in convection-permitting simulations. Earth and Space Science, 6, 818–835. https://doi.org/10.1029/2018EA000491.

Finney, DL , Marsham, JH , Jackson, LS et al. (6 more authors). (2019): Implications of improved representation of convection for the East Africa water budget using a convection-permitting model. Journal of Climate, 32,7, 2109-2129. https://doi.org/10.1175/JCLI-D-18-0387.1.

Finney, D. L., Marsham, J. H., Rowell, D. P., Kendon, E. J., Tucker, S. O., Stratton, R. A., & Jackson, L. S. (2020) Effects of explicit convection on future projections of mesoscale circulations, rainfall, and rainfall extremes over Eastern Africa. J. Climate, 33, 2701–2718, https://doi.org/10.1175/JCLI-D-19-0328.1.

Finney, DL , Marsham, JH , Wilkinson, JM et al (6 more authors). (2020b). African lightning and its relation to rainfall and climate change in a convection-permitting model. Geophysical Research Letters, 47, 23, https://doi.org/10.1029/2020GL088163.

Hart, N. C. G., Washington, R., & Stratton, R. A. (2018). Stronger local overturning in convective-permitting regional climate model improves simulation of the subtropical annual cycle. Geophysical Research Letters, 45, 11, 334–11,342. https://doi.org/10.1029/2018GL079563.

Jackson, L. S., Keane, R. J, Finney, D. L., Marsham, J. H., Parker, D. J., Senior, C. A., & Stratton, R. A. (2019). Regional differences in the response of rainfall to convectively coupled Kelvin waves over tropical Africa. J. Climate, 32, 8143–8165, https://doi.org/10.1175/JCLI-D-19-0014.1.

Jackson, L. S., Finney, D. L., Kendon, E. J., Marsham, J. H., Parker, D. J., Stratton, R. A., Tomassini, L., & Tucker, S. (2020). The Effect of Explicit Convection on Couplings between Rainfall, Humidity, and Ascent over Africa under Climate Change, J. Climate, 33, 19, 8315-8337. https://doi.org/10.1175/JCLI-D-19-0322.1.

Kendon, EJ., Stratton, RA., Tucker, S. (4 more authors). (2019). Enhanced future changes in wet and dry extremes over Africa at convection-permitting scale. Nature Communications, 10 1794. https://doi.org/10.1038/s41467-019-09776-9.

Misiani, H., Finney, D., Segele, Z.T., Marsham, J.H. Tadege, A., Artan, G., Athery, A. (2020). Circulation patterns associated with current and future rainfall over Ethiopia and South Sudan from a convection-permitting model. Atmosphere. Accepted

Mittal, N., Rowell, D.P., Dougill, A.J., Becker, B., Marsham, J.H., Bore, J., Tallontire, A., Vincent, K., Mkwambisi, D., Sang, J. 2020. Tailored climate projections to assess site-specific vulnerability of tea production. In review

Stratton R.A., C. A. Senior, S. B. Vosper, S. S. Folwell, I. A. Boutle, P. D. Earnshaw, E. Kendon, A. P. Lock, A. Malcolm, J. Manners, C. J. Morcrette, C. Short, A. J. Stirling, C. M. Taylor, S. Tucker, S. Webster & Wilkinson, J. M. (2018). A pan Africa Convection permitting regional climate simulation of the Met Office Unified Model: CP4-Africa. J. Climate https://journals.ametsoc.org/doi/pdf/10.1175/JCLI-D-17-0503.1.

Wainwright, C.M., Marsham, J.H., Rowell, D.P., Finney, D.L. and Black, E. (2020). Future changes in seasonality in Eastern Africa from regional simulations with explicit and parametrised convection. J.Climate, 1, 1-54. https://doi.org/10.1175/JCLI-D-20-0450.1

B. Other References

Finney, DL , Doherty, RM, Wild, O et al. (2 more authors). (2016). Response of lightning NOx emissions and ozone production to climate change: Insights from the Atmospheric Chemistry and Climate Model Intercomparison Project. Geophys. Res. Lett. 43 (10). pp. 5492-5500. https://doi.org/10.1002/2016GL068825.

Hanley, K.E., Plant, R.S., Stein, T.H.M., Hogan, R.J., Nicol, J.C., Lean, H.W., Halliwell, C. &Clark, P.A. (2015). Mixing-length controls on high-resolution simulations of convective storms. Q.J.R. Meteorol. Soc., 141: 272-284. doi:10.1002/qj.2356.

Hook, S., R. C. Wilson, S. MacCallum, & C. J. Merchant. (2012). Global Climate Lake surface temperature in "State of the climate in 2011". Bull. Amer. Meteorol. Soc., 93, 18-19. https://doi.org/10.1175/2012BAMSStateoftheClimate.1

Kendon, E. J. et al. (2017). Do convection-permitting regional climate models improve projections of future precipitation change? Bull. Am. Meteorol. Soc. 98, 79–93.10.

Marsham, J. H., N. Dixon, L. Garcia-Carreras, G. M. S. Lister, D. J. Parker, P. Knippertz, & C. Birch. (2013). The role of moist convection in the West African monsoon system – insights from continental-scale convection-permitting simulations, Geophys. Res. Lett., https://doi.org/10.1002/grl.50347.

Mizielinski, M. S. et al. (2014) High resolution global climate modelling; theUPSCALE project, a large simulation campaign. Geosci. Model Devel. 7, 1629–1640. https://doi.org/10.5194/gmdd-7-563-2014.

Pichelli E, Coppola E, Sobolowski S, Ban N, Giorgi F, Stocchi P, Alias A, Belusic D, Berthou S, Caillaud C, M Cardoso R, Chan S, et al. (2020). Themfirst multi-model ensemble of regional climate simulations at kilometer-scale resolution part 2: future precipitation projections. Climate Dynamics, under review.

Poulter, B., & Coauthors.(2015). Plant functional type classification for earth system models: results from the European Space Agency's Land Cover Climate Change Initiative.Geosci. Model Dev., 8, 2315-2328. http://hdl.handle.net/11858/00-001M-0000-0028-41D5-6.

Prein, A. F et al. (2017). Challenges and Advances in Convection-Permitting Climate Modeling, Bull. Am. Meteorol. Soc. 98, 1027 – 1030. https://doi.org/10.1175/BAMS-D-16-0263.1

Reynolds, R., T. Smith, C. Liu, B. Chelton, K. Casey, and M. Schlax, 2007: Daily high-resolution-blended analyses for sea surface temperature. J. Clim., 20, 5473-5496, https://doi.org/10.1175/2007JCLI1824.1.

Walters, D et al., 2018: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES global land 7.0 configurations. Geosci. Model Dev. https://doi.org/10.5194/gmd-12-1909-2019.

Weedon, G. P., G. Balsamo, N. Bellouin, S.~Gomes, M. J. Best, and P. Viterbo, 2014: The WFDEI meteorological forcing dataset: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. Water Resources Research, 50 (9), 7505-7514, https://doi.org/10.1002/2014WR015638.

For more information on IMPALA visit: https://futureclimateafrica.org/project/impala/,

or contact info@futureclimateafrica.org

www.futureclimateafrica.org

Twitter: future_climate

Disclaimer

Funded by