Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades der
Bayerischen Julius-Maximilians-Universität
Würzburg

Impacts of climate variability and change on Maize (*Zea mays*) production in tropical Africa

vorgelegt von
Freddy Fefe Bangelesa
aus Kinshasa

Würzburg 2021

Eingereicht am:	29. June 2021
1. Gutachter:	Prof. Dr. Heiko Paeth
2. Gutachterin:	Prof. Dr. Barbara Sponholz
der Dissertation	
Mentor*innen:	Prof. Dr. Heiko Paeth
	Prof. Dr. Christopher Conrad
	Prof. Dr. Birgit Terhorst
1. Prüfer:	Prof. Dr. Heiko Paeth
2. Prüferin:	Prof. Dr. Barbara Sponholz
der mündlichen Prüfung	
Tag der mündlichen Prüfung:	17. Dezember 2021
Doktorurkunde ausgehändigt am:	

Abstract

Climate change is undeniable and constitutes one of the major threats of the 21st century. It impacts sectors of our society, usually negatively, and is likely to worsen towards the middle and end of the century. The agricultural sector is of particular concern, for it is the primary source of food and is strongly dependent on the weather. Considerable attention has been given to the impact of climate change on African agriculture because of the continent's high vulnerability, which is mainly due to its low adaptation capacity. Several studies have been implemented to evaluate the impact of climate change on this continent. The results are sometimes controversial since the studies are based on different approaches, climate models and crop yield datasets. This study attempts to contribute substantially to this large topic by suggesting specific types of climate predictors. The study focuses on tropical Africa and its maize yield. Maize is considered to be the most important crop in this region. To estimate the effect of climate change on maize yield, the study began by developing a robust cross-validated multiple linear regression model, which related climate predictors and maize yield. This statistical transfer function is reputed to be less prone to overfitting and multicollinearity problems. It is capable of selecting robust predictors, which have a physical meaning. Therefore, the study combined: large-scale predictors, which were derived from the principal component analysis of the monthly precipitation and temperature; traditional local-scale predictors, mainly, the mean precipitation, mean temperature, maximum temperature and minimum temperature; and the Water Requirement Satisfaction Index (WRSI), derived from the specific crop (maize) water balance model. The projected maize-yield change is forced by a regional climate model (RCM) REMO under two emission scenarios: high emission scenario (RCP8.5) and mid-range emission scenario (RCP4.5). The different effects of these groups of predictors in projecting the future maize-yield changes were also assessed. Furthermore, the study analysed the impact of climate change on the global WRSI. The results indicate that almost 27 % of the interannual variability of maize production of the entire region is explained by climate variables. The influence of climate predictors on

maize-yield production is more pronounced in West Africa, reaching 55 % in some areas. The model projection indicates that the maize yield in the entire region is expected to decrease by the middle of the century under an RCP8.5 emission scenario, and from the middle of the century to the end of the century, the production will slightly recover but will remain negative (around -10 %). However, in some regions of East Africa, a slight increase in maize yield is expected. The maize-yield projection under RCP4.5 remains relatively unchanged compared to the baseline period (1982-2016). The results further indicate that large-scale predictors are the most critical drivers of the global year-to-year maize-yield variability, and ENSO – which is highly correlated with the most important predictor (PC2) – seems to be the physical process underlying this variability. The effects of local predictors are more pronounced in the eastern parts of the region. The impact of the future climate change on WRSI reveals that the availability of maize water is expected to decrease everywhere, except in some parts of eastern Africa.

Zusammenfassung

Weil die Folgen des Klimawandels die Lebensgrundlagen aller Lebewesen beeinträchtigen, ist der Klimawandel ein sehr relevantes Thema des 21. Jahrhunderts. Seine negativen Effekte betreffen bereits viele Sektoren unserer Gesellschaft und die Prognosen zeigen, dass sich die Auswirkungen des Klimawandels Mitte und Ende dieses Jahrhunderts verschärfen werden. Die Landwirtschaft ist besonders betroffen, denn sie ist sehr abhängig vom Klima. Da die Landwirtschaft als Hauptnahrungsquelle der Menschen gilt, ist es erforderlich sich mit den Problemen des Klimawandels rechtzeitig zu beschäftigen, um in der Zukunft die Ernährung der Menschheit gewährleisten zu können. Viele Forscher beschäftigen sich mit den Folgen des Klimawandels in der Landwirtschaft. Besonders in Afrika wurde viel geforscht, weil die Landwirtschaft in Afrika sich technisch schlecht anpassen kann, um die Schwierigkeiten, die mit dem Klimawandel einhergehen, zu überwinden. Mehrere Studien wurden durchgeführt, um die Auswirkungen des Klimawandels in Afrika zu bewerten. Aufgrund der unterschiedlichen verwendeten statistischen Methoden, Modellierungen der Umweltprozesse oder Ertragsdaten sind die Ergebnisse teilweise kontrovers. Diese Studie versucht, einen wesentlichen Beitrag zum Einfluss des Klimawandels auf die Landwirtschaft in Westafrika zu leisten, indem sie spezifische Methoden vorschlägt, um das Klima der Zukunft projizieren zu können. Diese Studie behandelt Maiserträge in den Tropen Afrikas, da Mais dort die wichtigste Nutzpflanze ist. Um die Auswirkungen des Klimawandels auf den Maisertrag abzuschätzen, wurde ein Regressionsmodell (aus dem Englischen: robust cross-validated multiple) entwickelt, das Klimaprädiktoren und Maiserträge koppelt. Diese entwickelte statistische Übertragungsfunktion ist zuverlässiger bei Schwierigkeiten mit der Überanpassung und der Multikollinearität. Außerdem ist sie auch in der Lage robuste Prädiktoren mit physikalischer Bedeutung auszuwählen. Deshalb wurden in der Studie großräumige und lokale Prädiktoren kombiniert. Erstere entstammen der Analyse der Komponenten des monatlichen Niederschlags und der Temperatur, letztere basieren basieren auf den mittleren und Extremtemperaturen sowie dem mittleren Niederschlag. Zusätzlich zu den Prädiktoren

wurde ein Index der Wasserbedarfsdeckung (Water Requirement Satisfaction Index, WRSI) verwendet, der auf einem Wasserhaushaltsmodell der Nutzpflanzen basiert. Die erwartete Mais-Ertragsänderung wird mithilfe eines regionalen Klimamodells (RCM) REMO für die Emissionsszenarien RCP8.5 und RCP4.5 simuliert. Die einzelnen Effekte der Prädiktoren-Gruppen bei der Prognose der zukünftigen Mais-Ertragsänderungen wurden ebenfalls bewertet. Darüber hinaus analysierte die Studie die Auswirkungen des Klimawandels auf den WSRI. Durchschnittlich zeigen die Ergebnisse eine jährliche Maisproduktionsänderung von ca. 27 % in der gesamten Region. Diese Änderung, die in Westafrika mit ca. 55 % stärker ausgeprägt ist, ist eine Folge des Klimawandels. Die Simulationen des Modells anhand von RCP8.5-Emissionsszenario zeigen auch, dass der Maisertrag der gesamten Region voraussichtlich bis Mitte des Jahrhunderts abnehmen wird. Danach findet eine geringe Ertragserhöhung statt, die jedoch um ca. 10 % unter der ursprünglichen Menge liegt. Im Gegensatz zu Westafrika wird in einigen Regionen Ostafrikas wird ein leichter Anstieg des Maisertrags simuliert. Die Mais-Ertragsprognose für die gesamte Region mittels RCP4.5 bleibt relativ unverändert im Vergleich zum ursprünglichen Ertrag. Die Ergebnisse zeigen weiterhin, dass die großräumigen Prädiktoren die wichtigste Rolle bei den globalen jährlichen Maisertragsschwankungen spielen. ENSO ist stark mit dem wichtigsten Prädiktor korreliert, was auf den physikalischen Prozess hinweist, der diese Ertragsänderung erklärt. Die Relevanz der lokalen Prädiktoren ist in den östlichen Regionen Afrikas stärker ausgeprägt. Sie beeinflussen den WRSI, sodass der Maisertrag im Verhältnis zur Wasserverfügbarkeit voraussichtlich überall abnehmen wird. Ausgenommen sind einigen Regionen Ostafrikas.

Résumé

A ce jour, il est clairement établi que le changement climatique est un phénomène évident sans équivoque qui constitue une problématique majeure du 21ème siècle étant donné ses revers sur la vie. En effet, il affecte négativement et sensiblement plusieurs secteurs de notre société et il en sera davantage, et selon plusieurs projections. De ce fait, une emphase sur les effets du changement climatique sur le secteur agricole, source principale de la nourriture, mérite d'être porté car celui-ci est fortement dépendant des variabilités climatiques. Plusieurs chercheurs se sont penchés sur la question, et une attention considérable a été accordée aux impacts du changement climatique sur l'agriculture en Afrique, en raison de la grande vulnérabilité du continent, due principalement à sa faible capacitée d'adaptation. Cependant les résultats s'avèrent parfois controversés à raison des différences entre les approches statistiques, les modèles climatiques, et les données de production agricoles utilisés. Cette étude tente d'apporter une contribution substantielle à ce vaste sujet, en suggérant des prédicteurs climatiques spécifiques. Nous nous sommes concentrés sur l'Afrique tropicale avec un focus sur la culture du maïs, considérée comme la plus importante de la région. Pour quantifier l'impact du changement climatique sur le rendement en maïs en Afrique tropicale, nous avons développé une régression linéaire multiple à validation croisée, qui connecte les prédicteurs climatiques et à la production annuelle du maïs. Le modèle ainsi développé est moins affecté par le surajustement statistique « overfitting » et la multi-colinéarité, et est capable d'identifier des prédicteurs robustes ayant une interprétation physique. Nous avons donc combiné : des prédicteurs à grande échelle, obtenus après l'analyse en composantes principales des précipitations et des températures mensuelles ; des prédicteurs traditionnels à l'échelle locale, principalement la précipitation moyenne, la température moyenne, la température maximale et la température minimale ; et les indices de satisfaction des besoins en eau, dérivant du modèle de bilan hydrique des cultures. La projection de la production du maïs a été simulée à l'aide d'un modèle climatique régional (MCR) REMO sous le scénario d'émission RCP8.5 et RCP4.5. Les effets séparés de groupes de prédicteurs sont également évalués. En plus,

nous avons analysé l'impact du changement climatique sur l'indice annuelle de satisfaction en besoin en eau. Les résultats de notre étude montrent que les variables climatiques expliquent en moyenne 27 % de la variabilité interannuelle de la production de maïs dans la région d'étude entière. Cette une influence des prédicteurs climatiques sur la production de maïs est plus prononcée en Afrique de l'Ouest, expliquant plus ou moins 55 % de sa variabilité. Les projections du model montre également, selon le scénario d'émission RCP8.5, que la production de maïs devrait diminuer d'ici 2050 et qu'une légère amélioration s'observerait après 2050 jusqu'à la fin du siècle, mais que cette production resterait toujours en baisse par rapport à la période de référence (environ -10 %) dans la région entière. Toutefois, dans certaines régions d'Afrique de l'Est, une légère augmentation en production du maïs est projetée tout le long de la période considérée pour la simulation. Sous le scenario d'émission RCP4.5, le changement en production du maïs par rapport à la période de référence (1982-2016) reste relativement inchangé dans toute la région. Les résultats indiquent en outre, que les prédicteurs à grande échelle contribuent plus à la variabilité interannuelle du rendement en maïs et ENSO – qui est fortement corrélé avec le prédicteur le plus important (PC2) - semble être le processus physique majeur expliquant cette variabilité. Les effets des prédicteurs locaux sont plus ressentis dans les parties orientales de la région. L'impact du changement climatique sur l'indice annuelle de satisfaction en besoin en eau révèle que la disponibilité en eau pour le maïs diminuerait partout dans la région, sauf dans certaines parties de l'Afrique de l'Est.

Contents

Abstract **i**

Zusammenfassung **iii**

Résume **v**

List of Figures **xvii**

List of Tables **xix**

List of acronyms **xix**

1 Introduction **1**

1.1 Motivation and background 1

1.2 Problem statement 4

1.3 Research Objectives 5

1.4 Structure of the thesis 5

2 Climate change impacts on agriculture in tropical Africa **7**

2.1 Climate change emission scenarios 7

2.2 Climate model downscaling 9

2.3 Climate model bias correction 10

2.4 Climate change in tropical Africa 11
2.5 Methods of assessing climate change impacts on crop yields 13
2.6 Impact of climate change on crop yield in tropical Africa 14

3 Study area **17**
3.1 Study location, relief and hydrography 17
3.2 Population and agriculture in tropical Africa 19
3.3 Climate of tropical Africa 20
3.4 Vegetation and land cover changes 23
3.5 Geology, geomorphology and soil of tropical Africa 24

4 Database **27**
4.1 Observational climate data 27
4.1.1 Climate Research Unit (CRU) data 28
4.1.2 Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4) 30
4.1.3 Reanalysis data (ERA5) 32
4.2 Climate model data 34
4.2.1 ECHAM6 35
4.2.2 REMO 37
4.3 Climate teleconnection patterns and indices 40
4.4 Crop yield dataset 44

5 Statistical methods and data pre-processing **48**
5.1 Data pre-processing 48
5.1.1 Spatial Interpolation 48
5.1.2 Quantile mapping bias correction 49

5.2 Statistical methods 57
5.2.1 Numerical summary measures 58
5.2.2 Person correlation 60
5.2.3 Linear regression and trend removal 60
5.2.4 Non-linear trend and abrupt change analysis 62
5.2.5 EOF analysis 64

6 Specific crop (Maize) water balance model estimation 69
6.1 Reference evapotranspiration 71
6.2 Crop water requirement 74
6.3 Actual evapotranspiration 76
6.4 WRSI 78

7 Development of the statistical crop yield model 79
7.1 Cross-validated multiple regression 80
7.2 Cross validation, bootstrap sampling and variable selection 81
7.3 Model weighting 83
7.3.1 Simple averaging 84
7.3.2 Adjusted coefficient of determination 84
7.3.3 The corrected Akaike Information Criteria (AICc) 85
7.3.4 Bayesian information criteria (BIC) 86
7.3.5 Bayesian moving average (BMA) 86
7.4 Prediction of the maize yield 87

8 Climate Impact Assessment 88
8.1 Statistical crop model for Maize in tropical Africa 88

8.1.1 Traditional and process-based Local Predictors 88
8.1.2 Large-scale climate predictors 93
8.2 Evaluation of present-day crop models 98
8.2.1 Impact of the group of predictors on the Maize yield model 98
8.2.2 Impact of multi-model averaging techniques on the Maize yield model 100
8.2.3 Model variable selection 102
8.3 Modelling Maize yield under the future climate change Condition 105
8.3.1 Projection of maize yield changes under climate change conditions 106
8.3.2 Multi-Model averaging techniques and the future Maize yield change 110
8.3.3 Separated effects of predictors/groups of predictors and future Maize yield Changes 110
8.4 Projected of WRSI under climate change conditions 111

9 Discussion and recommendations **114**

Bibliography **122**

Acknowledgment **153**

List of Figures

3.1 Location map of the study area, showing terrain (shading: altitude in meters above mean sea level), some water bodies and the country boundaries. The blue dashed lines indicate the regions where datasets (climate and crop yields) have been extracted 18

3.2 Map showing the Köppen climate classification of Africa (Peel et al., 2017). 22

3.3 Land cover/land use map of Africa (modified from Defourny et al., 2014). 24

3.4 Map of soil orders in Africa Source: FAO (2000). 26

4.1 Spatial pattern of the 1979–2018 climatology annual mean precipitation (left) and temperature (right) from the CRU dataset. 29

4.2 (a) Spatial climatology of mean SST; and (b) SST times series for the Pacific and Atlantic Oceans of the concerned region, defined by the dashed black line in the spatial pattern. The data come from ERSST.v4. 32

4.3 Climatology yearly means of selected variables: (a) total precipitation in mm; (b) mean temperature in $°C$; (c) 10-meter wind in m/s; (d) surface runoff in $mm/year$; (e) dew point temperature in $°C$; (f) surface pressure in Pa; (g) solar radiance $MJ.m.d^{-1}$; and (h) soil moisture in $mm/year$. 34

4.4 ECHAM6 SST from 1900 to 2020 37

4.5 Spatial climatology of selected variables from historical REMO (1980-2005): (a) total precipitation in mm; (b) mean temperature in $°C$; (c) 10-meter wind in m/s; (d) surface runoff in $mm/year$; (e) dew point temperature in $°C$; (f) surface pressure in Pa; (g) solar radiance $MJ.m.d^{-1}$; and (h) soil moisture in $mm/year$. 38

4.6 Climate indices and teleconnection patterns from 1948 to 2010. Source: https://psl.noaa.gov/data/climateindices/. 41

4.7 Crop yield times series data from 1960 to 2020 for selected African countries. Data source: FAOSTAT. 45

4.8 Gridded crop yield data: annual mean (left) and linear trend (right) from 1982 to 2011. 46

4.9 Cropping calendar for maize yield. 46

5.1 Cumulative density probability Function of the observed ERA5, corrected and uncorrected REMO. (Left) temperature, (right) precipitation. 52

5.2 Annual cycle of area averaged-temperature, precipitation, 10 m wind, dew point temperature, solar radiation, surface pressure, soil moisture and surface runoff. The green curve corresponds to the observed dataset (ERA5), blue is the uncorrected historical REMO, and the red is the corrected historical REMO (with empirical quantile mapping bias correction approach. 53

5.3 Spatial distribution of observation ERA5 (extreme left), historical raw simulated REMO (extreme right), and historical corrected REMO (middle) for mean temperature, mean precipitation, 10 m wind, dew point temperature, solar radiation, surface pressure, soil moisture and surface runoff. . 55

5.4 Spatial distribution of the difference between the historical REMO and observed ERA5 (REMO – ERA5) for mean temperature, mean precipitation, 10 m wind, dew point temperature, solar radiation, surface pressure, soil moisture and surface runoff. 57

5.5 Variance explained by each of the leading 50 EOFs of precipitation and temperture 67

6.1 Monthly variability of the soil moisture, reference evapotranspiration, actual evapotranspiration, precipitation and runoff. 70

6.2 Spatial pattern of the monthly reference evapotranspiration (1982–2016). 74

6.3 Crop coefficient curve for Maize (Allen, 1998) 75

6.4 Spatial pattern of the monthly actual evapotranspiration (1982-2016). . . 76

6.5 Inter-annual variability of WRSI in Johannesburg, Kinshasa, Porto Novo, Kigali and Kisangani. 78

8.1 Spatial distribution of the five-year means of WRSI-glo and linear trends (extreme top-right). 90

8.2 Spatial distribution of the WRSI calculated during different maize growing season stages: (a) initial; (b) developing; (c) mid-season; and (d) late season. 92

8.3 Linear correlation between maize yield and local climate variables: (a) maximum temperature, (b) minimum temperature, (c) mean precipitation, (d) mean temperature, (e) WRSI-lat, (f) WRSI-mid, (g) WRSI-glo, (h) WRSI-int, (i) WRSI-dev. Black dots indicate places where correlations are significant at the 5 % significance level. 93

8.4 Nine years moving average time series of the two leading principal component for September-October-November precipitation sums of CRU and REMO datasets, and their associated spatial pattern. PCs have been standardised relative to 1961–1991 climatology mean. 95

8.5 Nine years moving average times series of the two leading principal components for September-October-November temperature mean of CRU and REMO non-detrended datasets, and their associated spatial pattern. PCs have been standardised relative to 1961–1991 climatology. 96

8.6 Mean explained variance in % of the cross-validated regression at $RMSE_{CV}$ = Minimum. (a) Model with all predictors, (b) model with simple local predictors, (c) model with WRSI predictors, (d) model with large scale-predictors, and (e) the empirical probability density function of the explained variance (**M1** = model with all predictors; **M2** = model with WRSI predictors; **M3** = model with traditional predictors; **M4** = model with large-scale predictors). 99

8.7 Spatial distribution of the explained variance and the root mean squared errors of different multi-model averaging techniques (BMA = Bayesian Moving Averaging; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; AICc = corrected Bayesian Information criterion). 101

8.8 Empirical probability density function of the root mean squared error for different multi-model averaging techniques. 101

8.9 Mean number of cross-validated predictors at RMEcv = minimum. . . . 102

8.10 Number of times a given predictor is selected in the cross-validated regression, over 100 interations. 103

8.11 Number of times a given predictor yields a positive coefficient in the regression model, over 100 iterations. 104

8.12 Frequency of each predictor yielding a negative coefficient in the regression model over 100 iterations. 105

8.13 Projection of median maize yield changes under RCP8.5 and RCP4.5 climate scenarios and their corresponding confidential intervals at the 5 % significance level. The yield changes are calculated based on average maize yield production from 1982 to 2016. 106

8.14 Projection of median maize yield changes under RCP8.5 and RCP4.5 climate scenarios and their corresponding confidential intervals at the 5 % significance level for the central, eastern and western parts of the study area. The yield changes are calculated based on average maize yield production between 1982 and 2016. 108

8.15 Projected median of maize yield changes under RCP4.5 and RCP8.5 during the middle (2036–2066) and end (2071–2100) of the century. The uncertainties of the projection are expressed in the upper and lower 95 % percent confidence intervals. 109

8.16 Influence of multi-model averaging techniques on the projected maize yield changes. 110

8.17 Separated effect of a group predictors on future maize yield projection : (a) effects of group of predictors; (b) large scales predictors; (c) traditional local predictors; and (d) local WRSI predictors. 111

8.18 Spatial distribution of the difference between the historical WRSI-glo and future projected WRSI-glo under two emission scenarios (RCP4.5 and RCP8.5), by the middle and end of the century. 112

8.19 Projected WRSI under two climate emission scenarios (RCP8.5 and RCP4.5): (a) tropical Africa – whole study region; (b) central parts of the study region – the Congo Basin; (c) eastern parts of the study region; (d) western parts of the study region. 113

9.1 Vegetation crop land use and grass land use across RCPs. The gray areas denote the 95 percentiles scenario (source: Van Vuuren et al., 2011). . . 118

List of Tables

6.1 Classification of WRSI according to drought conditions and crop performance (FAO, 1986; Legesse, 2010) 70

6.2 Lengths of Maize development stages and climatic regions (days) (Allen et al.,1998). 73

8.1 Mean, maximum and minimum values of WRSI in selected African countries 91

8.2 Linear correlation between climate indices (Atlan_SST = Atlantic Seas Surface Temperature; Indi_SST = Indian Seas Surface Temperature; TNA = Tropical North Atlantic; TSA = Tropical South Atlantic; MEI = Multivariate ENSO Index; NAO = North Atlantic Oscillation; SAOD = South Atlantic Ocean Dipole; QBO = Quasi-Biannual Oscillation; IMI = Indian Monsoon Index; DMI = Indian Monsoon Dipole) and maize yield. * denotes the corrolation is significant at 5 5% significant level, and ** at 1 % significant level. 97

8.3 Nonlinear trend detection on the projected maize yield using Mann Kendall test 107

List of acronyms

AICc corrected Akaike Information Criteria

ARC2 African Rainfall Climatology version 2.0

ASTER Advanced Spaceborne Thermal Emission and Reflection Radiometer

BIC Bayesian Information Criteria

BMA Bayesian Moving Average

CFSR Climate Forecast System Reanalysis

CHIRP Climate Hazards Group InfraRed Precipitation

CMIP3 phase three of the Coupled Model Intercomparison Project

CMIP5 phase five of the Coupled Model Intercomparison Project

CORDEX COordinated Regional climate Downscaling EXperiment

CRU Climate Research Unit

ECHAM5 European Centre/Hamburg Model version five

ECHAM6 European Centre/Hamburg Model version six

ECMWF European Centre for Medium-Range Weather Forecasts

ENSO El Niño-Southern Oscillation

EOF Empirical Orthogonal Function

ERSST.v4 Extended Reconstructed Sea Surface Temperature Version 4

ESEM Earth System Model

FAO Food and Agriculture Organization

GCM General Circulation Model

GPCC Global Precipitation Climatology Centre

GPCP Global Precipitation Climatology Project

HadISST Hadley Centre Sea Surface Temperature

HAMOCC5 HAMburg Ocean Carbon Cycle model version 5

IOD Indian Ocean Dipole

IS92 IPCC Scenarios, 1992

ISM Idian Summer Monson

ITCZ Intertropical Convergence Zone

J JSBACH Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg

MEI Multivariate ENSO Index

MERRA-2 Modern-era Retrospective analysis for Research and Applications version 2

MOS Model Output Statistics

MPI Max Planck Institute

MPIOM Max Planck Institute Ocean Model

MSEcv Mean Square Error of the cross validation

MSEM Mean Squared Error of the model

MSWEP Multi-Source Weighted-Ensemble Precipitation

NAO North Atlantic Oscillation

OLS Ordinary Least Squared

PCA Principal Component Analysis

PCs Principal Components

QBO Quasi-biannual Oscillation

RCM Regional Climate Model

PCPs Representative Concentration Pathways

SA90 Scientific Assessment emission scenario, 1990

SAM Southern Annular Mode

SAOD South Atlantic Ocean Dipole

SRES Special Report on Emission Scenarios

SSA Sub-Saharan Africa

SSP5-8.5 Shared Socioeconomic Pathways 8.5

TAMSAT Tropical Applications of Meteorology using Satellite and ground-based observations

TNA Tropical North Atlantic

TNA Tropical Rainfall Measuring Mission

TSA Tropical South Atlantic

UDEL University of Delaware

WRSI Water Requirement Satisfaction Index

1. Introduction

1.1 Motivation and background

Climate change has become unequivocal, and its impacts on environmental, social and economic sectors are well established. The Intergovernmental Panel on Climate Change's (IPCC) Fourth Assessment Report has sufficiently documented – based on *in situ* measurements – the increasing trends in global air and ocean temperature, the rise of the global average sea level, the changes in precipitation, and the melting of ice and snow Bernstein et al. (2008). According to Hartmann et al. (2013), an increase in temperature of 0.72 $°C$ has been observed since 1950. Climate models projected an increase in global temperature ranging between 0.3 and 4.8 $°C$ by the end of the 21st century, according to emission scenarios (Kirtman et al., 2013).

However, the impacts of climate change are not heavily distributed and depend on the state of development of a given region. Africa is a continent that is highly vulnerable to the impacts of climate change, not for its exposure, but mostly for its low adaptation capacity (Hassan and Nhemachena, 2008; Barros et al., 2014; Sultan and Gaetani, 2016). Africa's economy depends on climate-sensitive sectors such as water resource management, agriculture, hydroelectricity, forestry and breeding (Masson-Delmotte et al., 2018), and its population is very poor. According to many studies, the continent has already been experiencing a trend of increasing temperature, with an accelerated rate over the last two decades and extreme precipitation events. The average warming increasing rate was estimated at 0.5 $°C$ per century (Jones and Moberg, 2003; Malhi and Wright, 2004; Trenberth et al., 2007; Christy et al., 2009). In opposite to the temperature, the direction of change in precipitation is still controversial, but, on average, precipitation has decreased during the 20th century. Frequent droughts were observed in southern Africa, Sahel and the Horn of Africa (Kotir, 2011). Severe droughts in West Africa have been reported since 1960 (Nicholson and Selato, 2000). According to Diem et al. (2014), tropical

Africa have been much affected by drought compared to other tropical regions. However, some areas such as East Africa make exceptions, where wetting precipitation regimes have been observed (Kotir, 2011).

The observed negative impacts of climate change are likely to propagate and worsen in the future. In Africa, climate models indicate an average temperature increase of 2 to 4.5 $^{\circ}C$ (1.7 times more than the global average) by the end of the century (Engelbrecht et al., 2015). The reduction of precipitation is expected in the south-east of Africa (Kotir, 2011). Hulme et al. (2001) reported that precipitation in West Africa is expected to decrease. However, these results may vary according to climate models and emission scenarios. As considered to be highly influenced climate change, especially in African countries, agriculture is likely to be drastically affected by the future climate trend.

Climate change poses challenges and opportunities for agricultural production and food security across the world (Tao et al., 2014): an opportunity for the production of C3 crops because they are likely to increase with additional CO_2 concentration in the atmosphere; a challenge for C4 crops such as maize because the effects of CO_2 are limited. Despite this controversy, the impact of climate change on crop yields in Africa is generally negative. Crop yield loss for the next decades across Africa have been predicted by several studies, but the magnitude of such a change remains uncertain (Roudier et al., 2011; Knox et al., 2012; Sultan and Gaetani, 2016). A meta-analysis study conducted by Knox et al. (2012) indicated that climate change was likely to reduce the yield production across Africa by 8 % by 2050. Compared to the rest of the world, agricultural production in Africa over the last 50 years has very much declined (Kotir, 2011). The impact of climate change on crop yield should be given more attention because the continent has already faced other challenges such as low irrigation, lack of fertilisers and low investments, all of which do not favour improving food security in the region.

Among crops, maize is the most cultivated crop and the largest contributor to the total value of staple crop production in Africa (Schlenker and Lobell, 2010). The production of maize in Africa depends on precipitation because farmers cannot afford fertilisers, implement irrigation schemes or adopt adequate technologies. As a result, maize production is vulnerable to climate change/variability (Shi and Tao, 2014), and its production is very low compared to other tropical regions. Cairns et al. (2013); Ngoma et al. (2019); Masasi and Ng'ombe (2019) have observed low production of maize in Sub-Saharan Africa (SSA) and attributed it to drought stress rather than poor irrigation schemes, pests, soil fertility, diseases or weeds. According to Amondo and Simtowe (2018), almost 70–80 % of maize loss in SSA is due to floods and drought. Numerous studies have projected a

negative impact of climate change on maize yield (Müller et al., 2011; Cairns et al., 2013; Tesfaye et al., 2015; Mulenga et al., 2017; Amadu et al., 2020). As predicted by the Paris Agreement, an increase in temperature of 2°C will shorten the maize growing season and decrease its production in Africa.

Different methods have been suggested to study the impact of climate change on crop production on various spatial scales. This includes statistical analysis based on historical climate and crop yield data (Lobell and Field, 2007; Schlenker and Roberts, 2009; Blanc, 2012; Blanc and Strobl, 2013) and process-based crop simulation (Alexandrov and Hoogenboom, 2000; Butt et al., 2005; Deryng et al., 2014; Rosenzweig et al., 2014; Falconnier et al., 2020). Process-based crop simulation provides satisfactory results when crop parameters are well-calibrated, but this is not the case in tropical Africa (Lobell and Burke, 2010). The statistical approach offers an opportunity to relate crop yield datasets to climate predictors and it usually outperforms the process-based model on larger spatial scales (Lobell and Burke, 2010). However, it cannot integrate adaptation options and requires a dense and good quality climate observational dataset. For its implementation, local climate variables, such as mean precipitation, mean temperature, maximum temperature and minimum temperature, are mostly used, as predictors. Still, they are often prone to bias, especially in tropical Africa, where stations are scarce and RCMs do not reproduce small-scale processes very well. To reduce uncertainties related to the poor local climate variable quality in Africa, Awoye et al. (2017) successfully introduced large-scale climate variables to model the impact of climate change on many crop yields in Benin. The results were highly satisfactory, because large-scale variables represent large-scale processes, which are less prone to bias. In the same optic of improving crop yield response to climate change/variability, recent studies have tried to combine the advantages of both modelling approaches (Roberts et al., 2013; Urban et al., 2015; Blanc, 2017; Roberts et al., 2017). This approach benefitted from the capabilities of processed-based models while preserving the advantages of statistical models (Blanc, 2017). Roberts et al. (2017) reported that the combined approach was substantially better than those produced by each model separately. These results have been confirmed by Blanc (2017) using a global crop yield dataset.

In this respect, the statistical transfer function developed in this study is based on the combination of predictors, which are on the cutting edge of statistical crop yield modelling: traditional local-scale predictors; large-scale predictors (Awoye et al., 2017); and processed-predictors, based on the Water Requirement Satisfaction Index (WRSI) of different growing seasons. WRSI is an indicator of specific crop performance (McNally et al.,

2015), generated from a crop water balance model. WRSI has been widely used by internal organisations, such as the European Commission's Joint Research Centre, the Food and Agriculture Organization (FAO), and the US Agency for International Development, to monitor food security and develop early warning systems in Africa (Tarnavsky et al., 2018). In addition to being used as simple variables in the statistical model, we also further analyse the impact of climate change on the annual WRSI to understand the future availability of maize water in topical Africa. To our knowledge, no previous studies have combined these types of predictors in modelling crop production.

1.2 Problem statement

The projected impacts of climate change on crop yields is severe in tropical Africa because of the conditions prevailing in tropical Africa, dominated by a high population growth rate, poverty, and a high prevalence of malnourishment. According to Ravallion (2012), the population of tropical Africa lives on less than $ 1.25 a day. The number of malnourished people has gone from 173 million in 1990–1992 to 213 million in 2014–2016, representing 23.3 % of its total population (Von Braun, 2007). Other international organisations, such as The Economic Research Service of the US Agency for International Development – as cited by Tian and Yu (2019) – approximate that 301 million people (31 %) living in tropical Africa were exposed to food insecurity. Women and children under the age of five years are most likely to have poor nutritional status. For example, in the Democratic Republic of the Congo, 43 % of the children under five are stunted (chronic malnutrition). The high prevalence of malnourishment could be explained by armed conflicts, land tenure insecurity, the affectation of land demand for non-agricultural use, poverty, natural disasters, water scarcity, low investment in agriculture, degraded land and limit agricultural infrastructure (Lynd and Woods, 2011; Schwan and Yu, 2018). The trend of agriculture production has almost quadrupled between 1961 and 2014. Still its impact on food security has been alleviated by high population growth, resulting in the decline of food production per capita. Many published papers have addressed the connection between climate and crop yield across tropical Africa, but these investigations are mostly limited to sub-national and national levels. There is little insight into the impacts of climate change on crop yield on a global level. Therefore, this study analyses how climate variables are related to maize crop yield and how this relationship will hold in the future. The results will be helpful in improving the resilience of the agriculture sector.

1.3 Research Objectives

The main objective of this research is to predict the impact of future climate change on maize yield in tropical Africa by using cross-regression, which relates present-day local (process-based and traditional) and large-scale predictors to the observed maize yield dataset. This research is possible only if two major hypothesis stands: (1) climate variability and changes are only responsible for the maize yield changes and all other improving factors are controlled after removing the linear trend from the crop and climate variables datasets; and (2) the causes of present-day climate variables in the simulated climate model do not change in the future. Specific Objectives assigned of this study are:

- To develop a statistical crop model that relates maize yield to climate variables
- To understand which group of variables – those related to large scale climate teleconnection, those related to traditional local variables (precipitation and temperature) or those related to the basic process-based crop models – best estimate the year-to-year maize yield variability in tropical Africa and how they are likely to influence the future maize yield changes
- To evaluate how different climate scenario influence the future maize yield change
- To evaluation a range of statistical model averagings the climate-maize yield relationship
- To quantify the impact of future climate change on the global WRSI

1.4 Structure of the thesis

This work is structured into nine chapters. The first chapter presents the motivation, background, and objectives of the research. The second chapter reviews the state-of-the-art of the impact of climate change on agriculture production, particularly in the context of tropical Africa. Relevant concepts related to climate change and agriculture interaction are discussed in this chapter. The third chapter describes the characteristics of the study areas. The fourth chapter presents the sources of both observational and modelled datasets used in this study. The fifth, the sixth and the seventh chapters deal with the mythology used to obtain the results. The eighth chapter presents the

major results of the study related to maize yield modelling. Finally, the discussion and recommendations are presented in the last chapter.

2. Climate change impacts on agriculture in tropical Africa

This chapter presents recent and relevant literature on the impact of climate change on crop yield and essential concepts related to the climate-agriculture nexus. We address climate emission scenarios in the first section. The second and third sections discuss climate model downscaling and bias correction. In the fourth section, we reviewed different methods of modelling the impact of climate change on crop yield. In the fifth and sixth sections, we present climate change and its impact on crop yield in Africa in general and tropical Africa in particular.

2.1 Climate change emission scenarios

Climate change scenarios describe different pathways of how the future climate will respond to the rate of economic growth, demographic changes, emissions of greenhouse gases, aerosols, environmental and technological conditions and energy (Van Vuuren et al., 2011). Emission scenarios provide ideas about the fingerprints of anthropogenic activities on climate change. Therefore, they serve as inputs into climate models to achieve radiative forcing pathways to project changes and better understand uncertainties in mitigation and adaptation efforts. There are many published climate scenarios from different scientists and organisations (Carter et al., 2007). Those developed or approved by the IPCC are commonly used and include the scientific assessment emission scenarios developed in 1990 (SA90), the six IPCC Scenarios developed in 1992 (IS92) (Change, 1992), the Special Report on Emission Scenarios (SRES) (Nakicenovic et al., 2000) and the representative concentration pathways (RCPs) (Van Vuuren et al., 2011).

Regarding the history of the climate emission scenarios approved by the IPCC, the SA90 was the first climate emission scenario implemented in 1990. It was based on four emis-

sions scenarios and later completed by IS92, which incorporates economic growth, population and technology uncertainties. Then came the SRES climate emission scenario, which improved the previous version by covering more representative drivers of future emissions, demographic development and economic and technological progress (Nakicenovic et al., 2000). Many modelling teams developped SERES, which is composed of 40 scenarios, grouped in four main storylines, A1, A2, B1 and B2. Each storyline describes a specific demographic, social, economic, technological and environmental pathway. The quantification of storylines allows scientists to estimate greenhouse gases. SERES scenarios were used in phase three of the coupled model intercomparison project (CMIP3).

The need to develop RCPs comes from the lack of policy actions to mitigate climate change in previous emissions scenarios. RCPs contain emission, concentration and land-use trajectories and have been used in phase five of the coupled model intercomparison project (CMIP5). They cover many scenarios – which can be updated at any time – published in peer-reviewed literature. These scenarios are assembled in four groups, each corresponding to a specific radiative forcing pathway: RCP 2.6, RCP 4.5, RCP 6.5 and RCP 8.5. RCP 2.6 has a peak in radiative forcing at almost 3 W/m^2 before the end of the century (2100) and declines to 2.6 by the end of the century (Van Vuuren et al., 2011). RCP 4.5 and RCP 6.5 are considered stabilisation scenarios. RCP 4.5 has a pathway of 4.5 $W/m^2$and is close to SRES B1 (Wise et al., 2009). RCP 6.5 has stabilisation pathway at 6 W/m^2 and is close to SRES B2 (Wise et al., 2009). Finally, RCP 8.5 is a very high scenario pathway stabilised at 8.5 W/m^2 and close to SRESA1FI (Riahi et al., 2007).

More recently, the new generations of climate emission scenarios have been released for phase six of the coupled model intercomparison project (CMIP6), named the shared socioeconomic pathways (SSPs)(Gidden et al., 2019). These scenarios explore a large number of end-of-century radiative forcing values, ranging from 1.9 to 8.5 W/m^2, and four of them are based on the forcing levels of 2.6, 4.5, 6.5 and 8.5 W/m^2, a kind of updated version of the RCPs. One scenario has the forcing level of 1.9 W/m^2, which corresponds to the objectives of the Paris Agreement on climate change. Finally, the overshoot scenario exhibits a forcing peak and a reduction at the end of the century (3.4 W/m^2). SSPs are suitable to analyse warming pathways by which realistic climate outcomes may be reached. Advanced details on how SSPs have been built are provided by Gidden et al. (2019).

2.2 Climate model downscaling

GCMs (General Circulation Models) have helped to understand the climate system for a long time and have proven the ability to capture certain aspects of the climate system, such as the long-term precipitation trend on a global scale (Fotso-Nguemo et al., 2017a). However, they are limited in representing regional- and local-scale climate processes because of their coarse spatial resolution. Therefore, they are not suitable for direct use in adaptation planning and climate impact studies. In Africa, for example, GCMs have failed to represent the African easterly jet (Cook, 1999) and tropical easterly jet (Nicholson and Grist, 2003), which are important climate features of tropical African regions. Scientists have been attempting to increase the spatial resolution of GCMs to simulate local and regional processes despite of the high computational of GCMs. Climate model downscaling techniques that bring complementary information to GCMs are applied to bridge this information gap. These techniques are twofold: statistical and dynamical.

Giorgi et al. (2001) identified two categories of climate dynamical downscaling methods: high-resolution atmospheric GCMs (McGregor, 2015) and Regional Climate Models (RCMs). High-resolution atmospheric GCMs use information derived only from the atmosphere. They are therefore considered as recent versions of GCMs. In the last three decades, scientists have been promoting RCMs because they are less computationally demanding. RCMs are run in a limited area and use the initial boundary condition and Sea Surface Temperature (SST) of either GCMs or global reanalysis (perfect boundary condition) (Giorgi and Gutowski Jr, 2015). They have shown remarkable ability in improving many climate variables than GCMs by capturing local, regional and mesoscale processes, such as sea breeze, mesoscale convective systems, tropical storms (Prein et al., 2015), soil moisture-precipitation interaction (Taylor et al., 2013) and tropical convection (Hart et al., 2018). Di Luca et al. (2015) and Rummukainen (2016) have provided further advantages of RCMs. In Africa, RCMs from the coordinated regional climate downscaling experiment (CORDEX) – the most important framework of RCMs – have demonstrated more improvement than GCMs in the simulation of the precipitation seasonal mean and annual cycle (Klutse et al., 2016) and extreme events (Dosio et al., 2015). In some cases the sign of future climate change of both RCMs and GCMs may be opposite (Teichmann et al., 2013; Dosio and Panitz, 2016). The climate variability within the nested domain of the RCMs can explain this. Dosio and Panitz (2016) observed a robust decrease of mean precipitation over the highlands of Ethiopia using CORDEX RCMs, while most of the GMCs indicated an increase. Furthermore, Dosio and Panitz (2016) reported oppo-

site signs between GCMs and CORDEX RCMs in projecting precipitation over Central Africa.

In contrast, statistical downscaling methods relate GCM predictors (i.e. mean sea level pressure field) to local predictands (*i.e.* temperature and precipitation) through a statistical transfer function (Maraun et al., 2019). The transfer function will then be applied to the outputs of GCMs (Maraun et al., 2019). There is no unanimity in classifying statistical downscaling methods. Schoof (2013) classified them into four groups: scaling methods, regression-based approaches, weather-pattern-based approaches and weather generators. Maraun et al. (2019) preferred three groups: the perfect prog, model output statistics (MOS) and the weather generator. The perfect prog model relates the observed predictand and predictors (from the reanalysis) in a regression model, which can be linear or non-linear, such as artificial neural networks. The regression model is then applied to the projection of the predictors as simulated by GCMs to generate future projections. MOS is similar to perfect prog, but the calibration is between the observed predictand and simulated predictors from climate models. Weather generators are stochastic models that explicitly model at least the marginal and temporal aspects of a meteorological variable, often even the relationships between a set of variables and sometimes also spatial dependence. Many authors have reviewed different statistical downscaling methods (Giorgi et al., 2001; Benestad, 2004; Benestad et al., 2016).

In general, statistical downscaling methods are preferred by many users because of their low computational requirements, simplification, effectiveness and fast calculation (Zhang et al., 2020). However, they are limited to a set of variables and require consistent meteorological observation. The advantage of dynamical downscaling methods is that they provide a good representation of local-scale climate dynamical processes (including feedbacks), especially when there are not enough observational data.

2.3 Climate model bias correction

Despite the effort of improving GCMs, RCMs are still prone to significant systematic biases (Christensen et al., 2008; Teutschbein and Seibert, 2010). It is also complicated to directly apply them in climate change impact studies, which mostly require absolute values of the variables. The source of these biases partially comes from the boundary conditions of GCMs (Rummukainen, 2010; Hall, 2014; Giorgi and Gutowski Jr, 2015), model parameterisations (*i.e.* convention parameterisations), land surface schemes and

the internal variability of RCMs (Teutschbein and Seibert, 2012). Fotso-Nguemo et al. (2017b) found a positive bias of the REMO in modelling precipitation over southern Cameroon, Equatorial Guinea and Gabon. The Swedish regional climate model also exhibited the dominance of systematic biases over Central Africa (Tamoffo et al., 2019a). In this respect, statistical post-processing, commonly known as bias correction, is therefore important. Statistical bias correction can also be applied directly to GCMs (Ahmed et al., 2013).

Bias correction is defined as a systematic difference between the modelled climate variables of the climate system and the corresponding observed ones (Maraun et al., 2017). It is based on the assumption of stationarity: the causes of bias remain invariant in the future (Maraun, 2016). Several bias corrections have been developed (Chen et al., 2011; Johnson and Sharma, 2012; Teutschbein and Seibert, 2012). They are categorised according to the statistics they attempt to improve: mean, standard deviation, long-term trend and/or distribution. Teutschbein and Seibert (2012) enumerated some of the methods of bias correction: linear scaling (Lenderink et al., 2007), local intensity scaling, power transformation (Leander and Buishand, 2007), variance scaling (Chen et al., 2011), distribution, quantile or probability mapping (Sennikovs and Bethers, 2009) and delta-change correction (Moore et al., 2008). MOS is also used in climate bias correction (Leander and Buishand, 2007; Dobler and Ahrens, 2008; Piani et al., 2010). The quantile mapping technique yields better results (Piani et al., 2010; Rojas et al., 2011). Quantile mapping is a widely used bias correction method, especially when dealing with stochastic variables, such as precipitation and solar radiation (Smith et al., 2014). Sangelantoni et al. (2019) and (Themeßl et al., 2012) demonstrated that the quantile mapping technique outperforms all variants of MOS in correcting the daily precipitation. Therefore, we adopt quantile mapping in our study.

2.4 Climate change in tropical Africa

Like in other regions of the world, tropical Africa is also vulnerable to the impacts of climate change. The temperature has increased in the past and is likely to increase in the whole region. Field and Barros (2014) observed an average temperature increase of 0.5 $°C$ and projected an increase exceeding 2 and 4 $°C$ by the end of the century in Africa. Porter et al. (2014) also estimated an increase in the temperature in tropical Africa of 1 to 4 $°C$ in the period 2081–2100 relative to 1986–2005. This is similar to Aloysius et al.

(2016) study, who projected an annual temperature increase of between 2 and 4 $°C$ at the end of the twenty-first century by using 25 CMIP5 GCMs. Fotso-Nguemo et al. (2017b) used REMO and projected an increase of 5 $°C$ in the northern part of Central Africa under the RCP 8.5 scenario. The temperature exhibited a southward gradient. Adhikari et al. (2018) projected an increase in median temperature of 1.5 to 5. 5 $°C$ by the end of the century for SSA. The most recent study based on CMIP6 indicated an increase of 3.5 $°C$ over the central east of Africa under the shared socioeconomic pathways 8.5 (SSP-8.5). Drought is also reported to increase in tropical Africa. Gidden et al. (2019) reported a drier condition in both the south and west of tropical Africa, while the east of tropical Africa is expected to be wetter. No significant changes in drought condition have been reported in the central region.

The direction of change in precipitation in response to climate change is still uncertain in SSA. A general increase in mean precipitation of between 2 and 20 % has been reported by Adhikari et al. (2018) over the whole region of tropical Africa, but with a significant level of uncertainty. At the subregional level, the majority of climate models have projected a substantial reduction in precipitation in the eastern and western Sahel regions and an increase in the east and south regions of Africa (Engelbrecht et al., 2015; Field and Barros, 2014). In eastern Africa, the precipitation is projected to increase by 5 to 75 % by the end of the century, as indicated by Sillmann et al. (2013). Shongwe et al. (2011) found that the wetter conditions in Eastern Africa were accompanied by less severe drought and intense precipitation during the rainy season. Persistent flood events have been projected in Eastern Africa (Shongwe et al., 2011; Field and Barros, 2014). Shongwe et al. (2011) analysed 12 GCMs and found an increase in precipitation intensity over East Africa. The magnitude of change between -9 to 27 % has been reported by Aloysius et al. (2016). In tropical West Africa, Sillmann et al. (2013) reported a decrease in the mean precipitation by 15 % , but with a significant uncertainty level. However, the frequency of extreme precipitation days is likely to increase in West Africa (Field and Barros, 2014; Vizy and Cook, 2012). The uncertainties of precipitation projection are more pronounced in Equatorial Africa because of missing observation data. Dunning et al. (2018) found no change in precipitation, like Haensler et al. (2013) study. However, Almazroui et al. (2020) indicated an increase in mean precipitation by the end of the century. Collins et al. (2013) also reported increased precipitation across central equatorial Africa during boreal winter, particularly in the east of the region. Fotso-Nguemo et al. (2017b) used 20 GCMs to assess the impact of climate change on extreme precipitation over Central Africa. They found an increasing trend of extreme precipitation in the Great Lakes areas,

northern Zambia, southern Chad and northern Cameroon. New et al. (2006) also reported an increase in extreme precipitation in the western and southern parts of tropical Central Africa. Sonkoué et al. (2019) found that the total wet-day precipitation has scientifically increased during boreal winter over Central Africa. They further reported the same trend in March–May. Haensler et al. (2013) combined global and regional datasets under different emission scenarios to assess the impact of climate change over Africa. They found that the intensity of extreme precipitation and the number of wet-day-spells during the rainy season are expected to change. Saeed et al. (2018) found that the length of the rainy season was significantly lower in tropical Africa and much more pronounced in the east, reaching 27 % .

2.5 Methods of assessing climate change impacts on crop yields

The impacts of climate change on crop yields are commonly assessed using three approaches: process-based models, the Ricardian model, and time series statistical models. The process-based approach uses the physiological processes (assimilation of carbon and nutrients, transpiration) – at every stage of crop growth – that govern crop growth to model the production of crops (Roudier et al., 2011; Lobell and Asseng, 2017). Additional information of the model comes from weather, crop management, soil properties and plant genetics. While Ricardian and time series statistical methods are based on empirical observations. Statistical time series models relate time series climate variables to time series observational crop yield data. Observational crop yield data often come from national statistical surveys, field measurements and surveys. The difference between the ricardian and time series statistical models is that ricardian model is cross-sectional and relates climate variables to the net income of farming systems.

The impact of climate variables as assessed by statistical and process-based methods does not substantially differ in most literature. Lobell and Asseng (2017) found almost no difference between both methods when using temperature to predict crop yield. Roudier et al. (2011), Knox et al. (2012) and Knox et al. (2016) also found that both methods yield approximately similar results. This could be explained by the fact that process-based models are developed from empirical observation, and climate variables used in statistical models are mostly those selected in process-based experiments. For example, it is more realistic to choose climate predictors that match the crop growing season and

connect to the crop growth process. A recent study has tried to use statistical models to emulate maize yield derived from process-based models (Blanc and Sultan, 2015). The study concluded that statistical crop models could predict the in- and out-of-sample crop yields generated by process-based models with high precision. The result was further confirmed by Blanc (2017), who even extended it to five crop models and four crops (rice, maize, soybean, and wheat).

However, despite this close connection between the two models, the scientific community prefers the statistical approach because it is cost-effective and easy to calibrate and validate (Roudier et al., 2011; Knox et al., 2016). Calibration of statistical models does not need additional fieldwork, for example. The use of statistical approaches is also facilitated by the increasing amount of weather and crop yield data. Furthermore, the statistical approach can be established directly on a large scale by using aggregate data (Roudier et al., 2011). The disadvantage of process-based models is that they are expensive to implement and require many parameters for their calibration, mostly in a specific region; thus, the parameters may differ for different varieties of the same crop. Additionally, it is challenging to model many other important processes, such as the influence of pesticides and insects. Consequently, extrapolation is not suitable. Like process-based models, statistical methods also have some limitations. They are criticised as being pessimistic by projecting large negative results (Schlenker and Roberts, 2009) and cannot explain the process underlying the crop yield. Furthermore, they do not consider climate change adaptation (Lobell et al., 2011) and assume that the relationship between past crop yields and climate variables will remain the same in the future (Lobell and Burke, 2009). Many other authers have described both the advantages and disadvantages of time series statistical and process-based models (Lobell and Field, 2007; Tao et al., 2008). A recent study conducted by Roberts et al. (2017) tried to combine the advantages of both methods. The result was substantially better than those produced by each model separately. This innovative approach opens doors to future research.

2.6 Impact of climate change on crop yield in tropical Africa

It is more than evident that climate change and variability have a significant impact on crop production. According to the IPCC report, climate change over the last 30 years has already reduced global crop production in the range of 3 to 15 %; cereals such as

rice and maize are mostly affected (Porter et al., 2014). Globally, crop production will likely to decrease, even for the most ecological climate scenario as agreed in the recent Paris Agreement (Challinor et al., 2014). In SSA, the impact of climate change on crop yields is generally negative (Lobell et al., 2008; Field and Barros, 2014; Conway et al., 2015; Sultan and Gaetani, 2016). However, some non cereal crops are likely to benefit from future climate change impacts. The production of cassava, for example, is expected to increase in both Central and Eastern Africa, according to Adejuwon (2006).

For the whole continent of Africa, Knox et al. (2012) estimated that future climate change would significantly reduce the production of maize, sorghum and millet, while cassava and rice would remind unchanged. Bean yields in most of the African regions are projected to decrease (Lobell et al., 2008; Thornton et al., 2009, 2011), and wheat yield is projected to decrease by 35 % in SSA (Nelson et al., 2009).

In tropical Africa, the western region seems to be most affected. The yield of maize is projected to decrease between 20 and 40 % by 2050 under the RCP 8.5 scenario (Challinor et al., 2014).Roudier et al. (2011) reported that crop yields would decrease, on average, by 11 % in the northern of West part of the tropical Africa. Sultan et al. (2014) also found that cereal yields would decrease by 10 % in the middle of the century. However, it is important to mention that, as the region is projected to receive abundant precipitation, an increase in some crop yields is expected. Adejuwon (2006) has already reported a possible increase of cereal and cassava yields in Nigeria by the end of 2050 and a flattening of the curve until 2100. According to some studies, the production of peanuts is also expected to increase(Tingem and Rivington, 2009; Dube et al., 2013). However, other scientists have also reported negative impacts of climate change on peanut production (Schlenker and Lobell, 2010; Lobell et al., 2008).

In East Africa, the impact is not always negative. In Kenya, for example, Rosenzweig et al. (2014) has projected changes in maize yield in the order of 5 %. This is partially explained by the warming of the highlands in Eastern Africa. In contrast, more recently, Davenport et al. (2018) found that maize yield will be reduced by 8 % . In the same vein, Adhikari et al. (2018) reported that the impact of climate change on food projection was largely negative in the east of tropical Africa; millet yield was projected to reduce by 72 % and maize, rice and soybean yield by almost 45 %.

3. Study area

This chapter describes the socio-economical and biophysical characteristics of the study area. This includes its location, relief, hydrology, population, agriculture, climate, vegetation, geology and soil.

3.1 Study location, relief and hydrography

This study was conducted in tropical Africa between 20° north and 20° south, and 20° west and 53° west (Figure 3.1). To better understand the characteristics of the study region. We spitted the entire region into three sub-regions: tropical East Africa, tropical Central Africa and tropical West Africa.

Tropical East Africa is characterised by elevation gradients and highly diverse topographical environments. In the same region, can be found a point of 153 m below the mean sea level (Lake Assail in Djibouti) and a point of 5,895 m above the mean sea level (Mount Kilimanjaro in Tanzania). A continuous north–south barrier of highlands emerges from the Red Sea to southern Tanzania. This includes the Ethiopian massif, considered the largest highland in Africa, with more than half of its land at 1500 m above sea level. In the southern of the Ethiopian massif lie two other highlands: the Kenyan highlands and the Rift mountains. The Kenyan highlands are in the extreme east and have an average elevation of 1500 to 2500 m. The Kenyan highlands are dominated by two mountains: Mount Kilimanjaro and Mount Kenya (5,199 m). To the west of the Kenyan highlands lie the Rift mountains, which run from western Uganda to southern Tanzania, with the Rwenzori mountains as the highest point. Between the Kenyan highlands and the Rwenzori mountains is located the Victoria Lake, the second largest lake in the world. Several other lakes are located in the region: Lake Abbe and Lake Abaya in the Ethiopian Rift valley; Lake Turkana, Lake Natron and Lake Eyasi in the eastern Rift; Lake Albert (5,300 km), Lake Edward, Lake Kivu, Lake Tanganyika (32,900 km) and Lake Rukwa in the

western Rift; and finally, Lake Malawi (or Nyassa, 29,600 km) to the south. In the north of tropical Eastern Africa, precisely in South Sudan, lies the White Nile river. Central tropical Africa is dominated by the Congo Basin lowlands. The average elevation of the region is 500 m. The region is surrounded by the East Africa highlands, the Cameroon highlands and the southern highlands. The Democratic Republic of the Congo is the largest country in these regions. Other countries can also be enumerated as part of the region: the Republic of Congo, Angola, Gabon, Central African Republic and Cameroon (Camberlin, 2018; Shanahan, 2018).

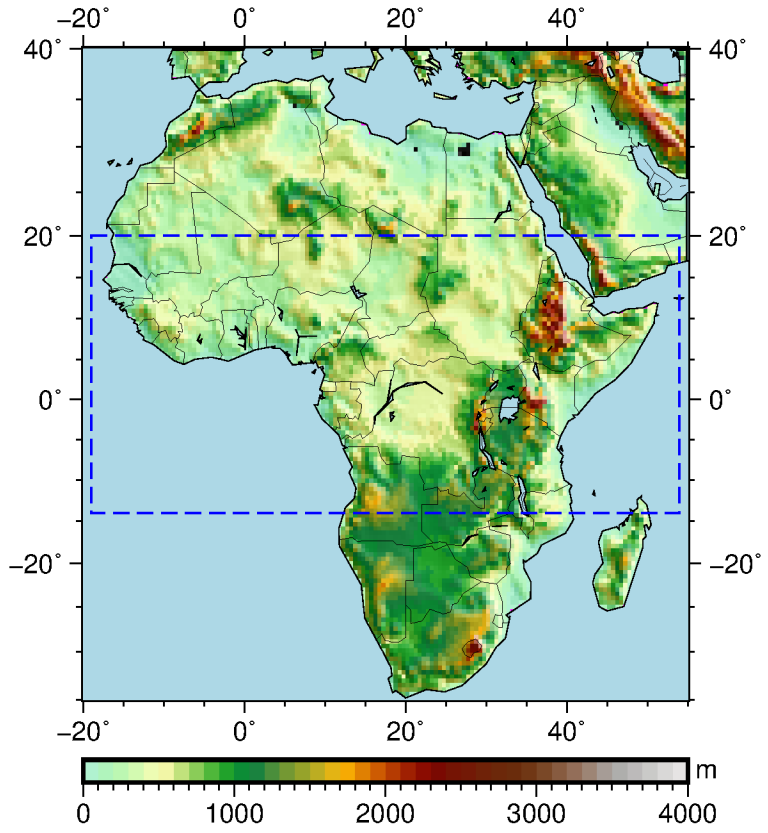

Figure 3.1: Location map of the study area, showing terrain (shading: altitude in meters above mean sea level), some water bodies and the country boundaries. The blue dashed lines indicate the regions where datasets (climate and crop yields) have been extracted

The relief of tropical West Africa is relatively high, compared to tropical Central Africa, but lower than Eastern Africa. The region is dominated by Guinea highlands, Jos plateau of northern Nigeria and Darfur in the Sudan. Further in the extreme south lie Adamawa

Mountains of Cameroon, which reach over 3000 m. Senegal and Niger are the longest rivers of the region. There are three other rivers: Bani, Ubangi, and Chari. Lake Chad is the most important lake in the region and has been, for a long time, affected by the impacts of climate change (Camberlin, 2018; Shanahan, 2018).

3.2 Population and agriculture in tropical Africa

As part of the SSA, Tropical Africa is among the poorest regions of the world, where almost 48 % of the estimated population live on less than $ 1.25 a day. In 2013, more than half of the world's extreme poor were in SSA (Anyanwu et al., 2017). This rate is alarming because two decades before (1990), this rate was only 15 %. According to the statistics of the World Bank, the number of people living in extreme poverty is projected to increase in the future if no serious measures are undertaken. Furthermore, the population growth, which is likely to influence the poverty rate in Africa, is also projected to double by 2050 (with 2004 taken as the reference) and reach two billion by the end of the century. To comply with this poverty rate and the population growth dynamic, crop production must be multiplied by a factor of five, as estimated by the UN (2012).

Considered as the principal source of income in SSA, almost 65 % of the population practice agriculture and contribute to the country's national gross domestic product (Ravallion, 2012). The agriculture sector is dominated by small-scale farmers who have limited financial resources.For example, in the east of tropical Africa, small-scale farmers contribute up to 90 % of the agricultural production (Salami et al., 2010). The crop yield has been increasing over the years, but there are still many challenges to address (Chauvin et al., 2012). Agriculture is still rainfed (sensitive to climate variability) and characterised by low productivity, lack of irrigation and limited access of fertilizers (Chauvin et al., 2012). Moreover, agriculture is not a priority for many countries' political agendas, resulting in less investment by the private and public sectors (Hoeffler, 2011). All these factors make it vulnerable to the impact of climate change. According to the World Bank, only 4 % of cultivated land is irrigated in SSA and less than 20 % of the fertilizers are applied in SSA (OECD/FAO, 2016). The impact of climate variability on agriculture is already perceptible. In 2015, the occurrence of El Niño was accompanied by extreme dry conditions in Eastern and Southern Africa, resulting in crop yield depletion, especially in Ethiopia, where the lowest annual precipitation was recorded since 1985.

Regarding types of crops, cereals (millet, maize, sorghum and wheat), as the main source of energy, are largely produced in SSA. Their production will rise by 41 Mt, relative to the 2013–2015 base period. The maize cropping system itself covers 27 % of the total cultivated land for cereal and represents 34 % of the total crop production in SSA. Furthermore, it has been considered as the most consumed cereal and cultivated staple food in humid tropical Africa because of its protein and calories. According to OECD/FAO (2016), the production of maize accounts for 40% and 70% of total cereal consumption and total maize demand respectively. In adequate precitation areas, maize produces larger yields than any other cereal. However, even in drier areas, where yields are more variable, maize is often planted for its higher quality compared to sorghum or millet and because it can be eaten green if necessary (Winch, 2006). The consumption of millet prevails in drier tropical Africa. Rice production has increased more than any cereal in SSA because of its ease of conservation and preparation and its ability to adapt to diverse environments: upland rice, lowland rainfed rice, mangrove swamp rice, floating rice and irrigated rice. Consumption and production are projected to increase by 18 % and 20 Mt, respectively (OECD/FAO, 2016). Wheat is the cereal produced least in SSA; it represents only 5% of cereal production. The statistics of other groups of foods and crops (roots, tubers, oilseeds, pulses, cotton, sugar, biofuels, meat, eggs, fish, dairy, fruit and beverage crops) are provided by (OECD/FAO, 2016).

3.3 Climate of tropical Africa

Unlike other regions of the world, such as North America, Europe and Asia, the temperature amplitude of the African tropical region, in general, is not fluctuating. The variability of the tropical African climate is explained by the movement of air masses, the relative stability and the difference in air moisture (Balek, 2011). These air masses intersect in the intertropical convergence zone (ITCZ), which is the most important weather feature of intra-annual precipitation fluctuation in tropical Africa. In Central Africa, the ITCZ reaches the northernmost position in July and the southernmost position in January. The phenomenon of air mass displacement is extensively described by Nicholson (2019).

The climate of West Africa and equatorial Africa is further influenced by the upper-level tropical easterly jet, the low-level westerly winds and the South and North African easterly jets. These three jets are associated with precipitation (Nicholson and Grist,

2003; Nicholson and Dezfuli, 2013; Nicholson, 2017). The West African monsoon is also a major climate feature that affects the precipitation in West Africa and equatorial Africa, where it coincides with the ITCZ (Nicholson, 2018). The effect of the West African monsoon does not reach East Africa because it is blocked by the East African highlands. According to Nicholson (2019), the precipitation in West Africa has a strong connection with the Atlantic Multi-decadal Oscillation. At the local level, precipitation in West and Central Africa is driven by mesoscale convective systems, larger cumulonimbus clouds that propagate westward. They explain up to 50 % of the precipitation in Africa, 90 % in West Africa and 85 % in equatorial Africa (Nesbitt et al., 2006). In East Africa, the general circulation is also influenced by the presence of the tropical easterly jet and the Somalian jet, and the effect of the ITCZ is diffused. In addition to the tropical jet, the Somalian and the low-level Turkana jets play an important role in precipitation variability (Nicholson, 2017). The El Niño-Southern Oscillation (ENSO) and the Indian Ocean dipole (IOD) also highly affect the region. Because of the topography of the region, the climate of the region is highly influenced by local features such as the Turkana jets, mountains and lakes.

According to the Köppen climate classification, considered the most widely used in the world, tropical Africa experiences four types of climate: tropical humid climate (Af and Am), tropical wet and dry climate (Aw), highland climate and dry climate. The tropical humid climate forms a belt less than 20 *degrees* wide across the Equator. It is characterised by the absence of winter and abundant precipitation accompanied by thunderstorms. The annual precipitation lies between 1750 and 2500 mm and reaches 3000 mm on the coast of the Gulf of Guinea (Nicholson and Dezfuli, 2013). The annual temperature varies between 10 and 25 $°C$, with a daily difference not exceeding 10 $°C$. There are two variants of this climate: the first is characterised by the abundant distribution of precipitation over the year and the second is influenced by the West African monsoon (Balek, 2011).

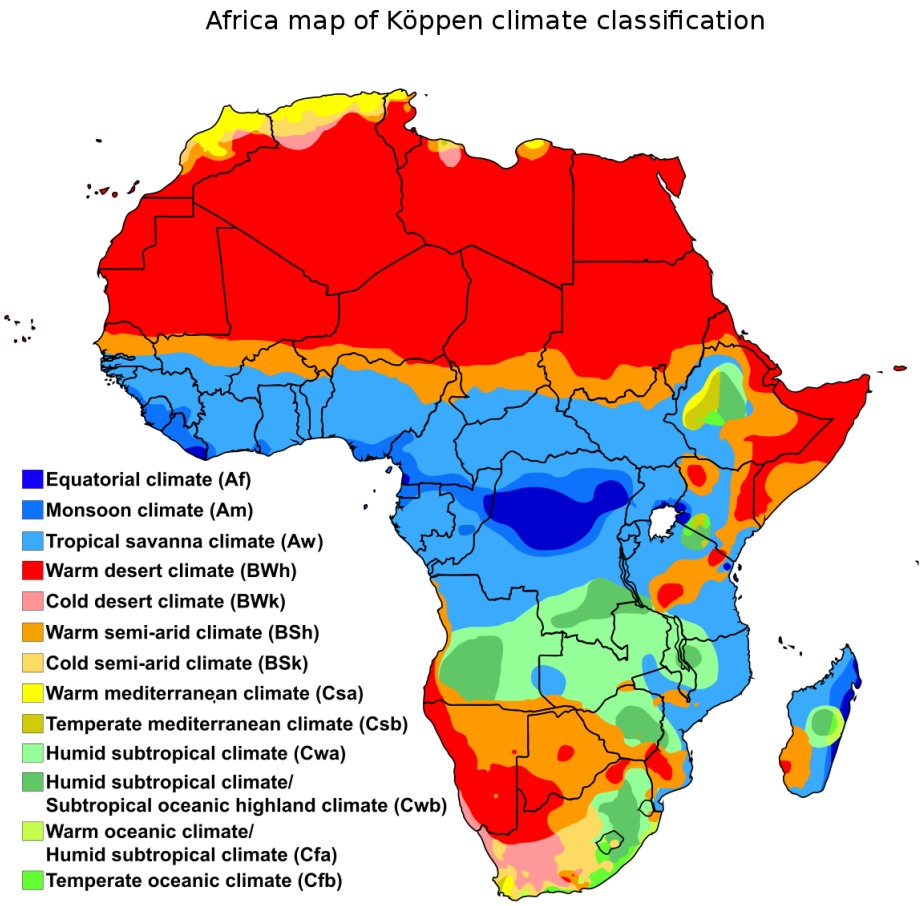

Figure 3.2: Map showing the Köppen climate classification of Africa (Peel et al., 2017).

The tropical wet and dry climate is distributed around the Equator, no more than 15 degrees to the north and reaching the Tropic of Capricorn to the south. The regions are characterised by the north–south movement of ITCZ, bringing precipitation during high sun. During low sun, the influence of the drier and the subtropical winds prevail. It is considered a transitional climate between the tropical humid climate and the dry climate. It has less annual precipitation than the tropical humid climate, ranging from 1000 to 1500 mm. The average annual temperature is usually higher than 30 $°C$. Warmer months in the north are March and April, while September, October and November are warmer months in the south (Balek, 2011).

The mountainous regions are characterised by highland climates. These are basically low-temperature variants of low elevations at a similar latitude, where the mean temper-

ature decreases by 0.65 to 0.95 $°C/100\ m$. Precipitation is generally heavier than the surrounding low land. This climate mostly prevails in the east of Africa (Balek, 2011).

The dry climate regions are characterised by water deficiencies. In the tropical region of Africa, there are two variants of dry climate: semi-arid climate and arid climate. The semi-arid climate is a transitional climate, between the desert climate and humid climate. The total annual precipitation ranges from 250 to 600 mm, while the arid climate is characterised by an annual precipitation of less than 250 mm and a temperature varying from 30 to 35 $°C$ (Balek, 2011). The arid climate is located in the coastal region of Eastern Africa, the hottest place in the world (Dallol in the Afar depression)(Pedgley, 1967).

3.4 Vegetation and land cover changes

According to Shant's classification of vegetation, the tropical region of Africa is subdivided into four biomes: mountain forest and grassland, equatorial rainforest, savanna (savanna, forest savanna and woodland savanna), dry forest and acacia savanna. The mountain forest and grassland colonise the tropical Eastern African highlands, above 1500 m. The equatorial forest dominates the Congo Basin in the low latitude where the temperature is high, approximately 24.5 $°C$, and the precipitation abundant. According to FAO (2016), almost 20 % of tropical forests are in Africa. The savanna, forest savanna and woodland savanna are the types of vegetation which surround the rainforest. It is which characterised by a tropical wet and dry climate. The dry forest, known also as the Miombo forest, located at the south of the Equator, precisely in the southern and south eastern parts. The acacia savanna and thorn desert are in areas where the precipitation is under 500 mm and the dry season is more extended (more than eight months).

The climate alone does not determine the distribution of tropical vegetation. Large-scale savannah are often found within a dense rainforest, mostly due to human activities. According to Staver et al. (2011), almost 36 % of tropical savannahs are located in areas where dense forests could form in normal conditions. Aleman et al. (2018) found that the tropical forest has lost almost 21 % of its cover since 1900. Compared to other tropical regions, the rate of deforestation is relatively low, probably because of the lack of infrastructure, low population density and political instability (Megevand, 2013). However, during the last two decades, the rate of deforestation has doubled (FAO-ITTO, 2011; Megevand, 2013), resulting in an increase of CO_2 in the atmosphere and modification of

the surface albedo, cloud physics and evapotranspiration (Bala et al., 2007; Lawrence and Vandecar, 2015). According to Nogherotto et al. (2013), deforestation may impact the West African monsoon and therefore change the general circulation over the Sahel region. If natural forest and savannah are the main carbon sinks, fire and deforestation constitute the main contributors of carbon sink losses (Bombelli et al., 2009). Agriculture is the second source of carbon sink losses In East Africa for example, cultivated areas increased by 28 % from 1990 to 2010 (Brink et al., 2014). The east and west regions of tropical Africa have lost 93 and 83 % of their cover since 1900 respectively, while forest expansion of about 1.4 % has been observed in Central Africa (Aleman et al., 2018). Similar results have been confirmed by other scientists who used different methodological approaches (Willcock et al., 2016).

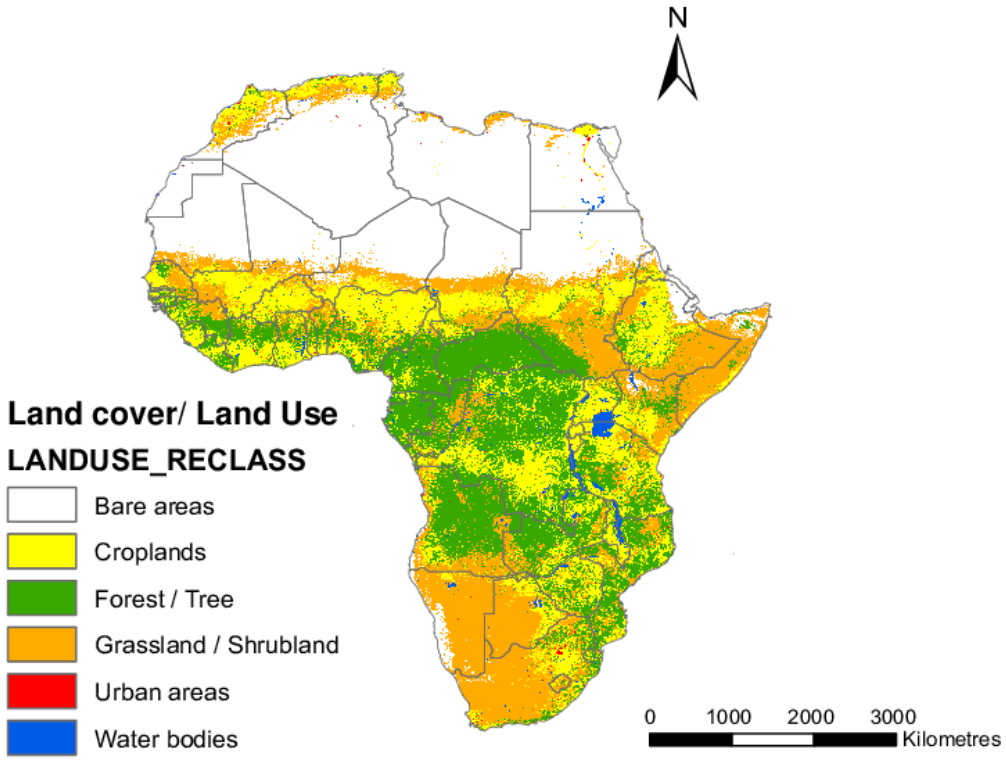

Figure 3.3: Land cover/land use map of Africa (modified from Defourny et al., 2014).

3.5 Geology, geomorphology and soil of tropical Africa

Tropical African geology and ores have recently been described by Milesi et al. (2006). In summary, African geology is dominated by the Archean craton formation, the Pale-

oproterozoic formation, the Mesoproterozoic to the early Neoproterozoic formation, the Neoproterozoic–Cambrian formation, Paleozoic–Mesozoic basins and the Cenozoic cover (Milesi et al., 2006). The Archean craton blocks cover a large part of West Africa (Equatorial Guinean, the Republic of Congo, the Gabon Massif and a large part of Cameroon), the Kasai-Lomami-Luanda corridor in the Democratic Republic of the Congo, the central shield of western Angola and the northern part of the Democratic Republic of the Congo. The Paleoproterozoic block covers the west of the Congo Craton, from Angola, and extends to the southwest of Cameroon. Another belt is located in central and northern Cameroon. Mesoproterozoic to early Neoproterozoic formations extend in the eastern part of Central Africa (east of the Democratic Republic of the Congo, Tanzania, Rwanda, Burundi and Uganda). The Neoproterozoic-Cambrian belt is located in the north, south, west and east of the Congo Craton. The Paleozoic–Mesozoic basin dominates the Central Africa region. The Cenozoic block covers the central basins of Congo and Angola, and some extend into Cameroon and the Virunga Massif (east of Democratic Republic of the Congo).

As with all tropical regions, the African tropical forest tends to be of low fertility, except for some parts that have developed in alluvial or volcanic parent materials (Milesi et al., 2006). The humid tropic soil is dominated by Ultisols, Oxisols and Alfisols (Figure 3.4). They are characterised by low cation exchange capacity, water holding capacity and plant nutrient reserves and are prone to erosion. In the tropical semi-arid region where Vertisols, Inceptisols, Alfisols and Aridisols prevail, the semi-arid tropical soil is prone to periodical drought.

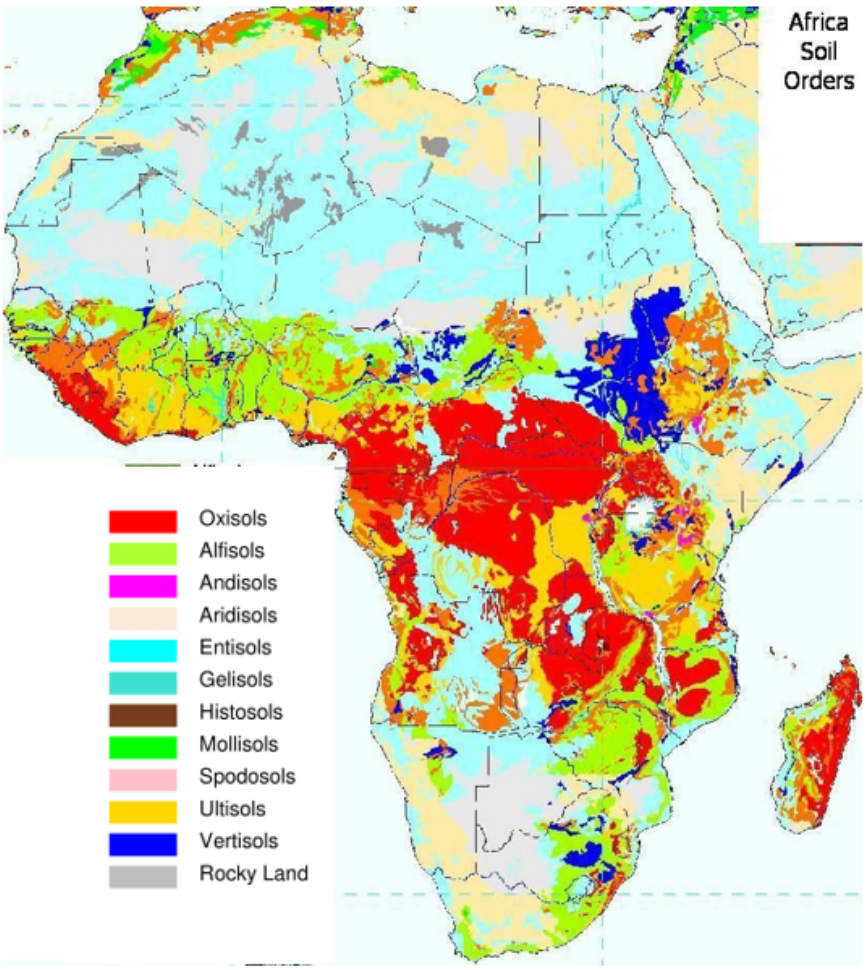

Figure 3.4: Map of soil orders in Africa Source: FAO (2000).

4. Database

This chapter describes the different types of datasets used to model the impact of climate change and variability of maize yield in tropical Africa. Two observational datasets were used: interpolated gridded dataset and reanalysis. The gridded dataset was chosen because of its long-term coverage, thus suitability to generate stable large-scale variables. The reanalysis dataset was used to complete the gridded dataset, lacking some local variables. The climate model datasets are the second type of dataset. The general circulation and the regional climate model were used to forecast the future maize yield change. Another type of data is climate indices; they were not put in the crop yield model but serve to understand the physical meaning of climate predictors. Finally, the historical crop yield dataset and its related crop calendar are presented.

4.1 Observational climate data

Climate observational databases are important to describe the main features of the present-day climate. They are also useful to validate and calibrate different models. Ideal observational climate datasets are those that come directly from ground stations. However, this is not always the case in Africa, where ground stations are scarce, inconsistent and have limited documentation and bad sharing policies, making climate data analysis difficult (Dinku et al., 2014). Therefore, data are directly collected from other network meteorological stations (provided by international organisations after some post-processing), satellite estimates, a combination of satellite and ground stations and reanalysis datasets.

Climate satellite-derived datasets have been collected over the last two decades. The most used in Africa are the African rainfall climatology version 2.0 (ARC2) with a spatial resolution of 0.1° and the Climate Hazards Group InfraRed Precipitation (CHIRP) with a spatial resolution of 0.05°. Some are combined with ground-based station obser-

vation, such as multi-source weighted-ensemble precipitation (MSWEP) and the tropical applications of meteorology using satellite and ground-based observations (TAMSAT). The shortcoming of satellite-derived datasets is that they cannot provide many climate variables. To fill this gap, reanalysis datasets are used. Some reanalysis datasets are commonly used in Africa: modern-era retrospective analysis for research and applications version 2 (MERRA-2), ERA-interim, ERA 5 and the climate forecast system reanalysis (CFSR) (Gebrechorkos et al., 2018). According to (Paeth et al., 2005), the validity of these datasets is questionable, and they are all prone to major errors. For example, almost all satellite devices overestimate precipitation in semi-dry regions. Nikulin et al. (2012) reported a significant dry bias of Tropical Rainfall Measuring Mission (TRMM) compared to gridded meteorological data. In contrast, the combined satellite and gouge data seem to be better than satellite data alone (Paeth et al., 2005). Still, they have not outperformed gridded methodological datasets, as demonstrated in West Africa (Ali et al., 2005).

4.1.1 Climate Research Unit (CRU) data

In this study, we chose to use a CRU dataset because of its availability, spatial and temporal resolution and long-term coverage. CRU data have been widely used in different research studies, including in crop yield modelling (Paeth et al., 2005; Deryng et al., 2014). The CRU data constitute a high-resolution grid and monthly data developed by the University of East Anglia. There are many versions of the CRU data, which have been resealed in 2000, 2004, 2005 and 2006. The first version was implemented by New et al. (2000). The last version, used in this study substantially improves the interpolation method and increases the variable count to 10 (Harris et al., 2020). Like its predecessors, it uses ground-based station data, which are further interpolated to a 0.5° regular grid using angular distance weighting interpolation. The dataset is made up of three types of variables: primary, secondary and derived variables. The primary variables are the mean 2 m temperature, the diurnal 2 m temperature range and the precipitation. They are all derived from station data. The secondary variables are those collected from a limited number of observational stations, thus completed by an empirical relation with primary variables, and these are vapor pressure, wet day and cloud cover. The final type of variable is entirely derived from the primary ones, and these are the minimum 2 m temperature, maximum 2 m temperature, frost days and potential evapotranspiration (Harris et al., 2020).

The weakness of CRU datasets is that the number of observed stations is not constant over the years and stations are quite sparse. For example, from 1970–1979 to 2000–2009, stations' numbers substantially decreased, especially in Central Africa. Despite this weakness, the cross-validation of CRU data with ground-based station observations is correlated for the temperature and highly correlated for the precipitation. The correlation is even stronger in high latitude regions, which have many stations. Independent studies have also evaluated the robustness of CRU datasets against station observations in some regions (Malsy et al., 2015; Shi et al., 2017). Shi et al. (2017) found that CRU data underestimate precipitation compared to the 29-point rain gauge records in headwater regions (China).

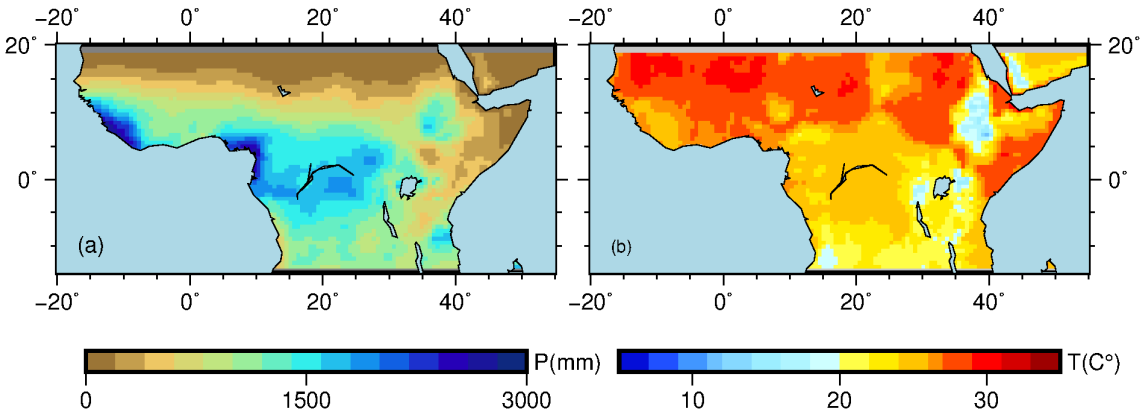

Figure 4.1: Spatial pattern of the 1979–2018 climatology annual mean precipitation (left) and temperature (right) from the CRU dataset.

Compared to other gridded datasets, the CRU temperature is highly correlated with the University of Delaware's 0.5 ° (UDEL) dataset and the reanalysis dataset JRA-55 from the Japan Meteorological Agency. There is a correlation coefficient of 0.97 between CRU and UDEL, and 0.99 between CRU and JRA-55. UDEL seems to be cooler than CRU in the early twentieth century. For the precipitation, CRU data have been compared to GPCC; they yield a correlation of 0.92, but GPCC seems wetter in the early twentieth century. In East Africa, Ongoma and Chen (2017) found that CRU outperformed GPCC in representing the annual precipitation cycle. Akinsanola et al. (2017) also reported good performance of CRU data together with TRMM and GPCC to assess precipitation over West Africa. The inter-annual anomaly of the three datasets showed a high correlation with ground observation. The previous versions also showed a satisfactory agreement be-

tween CRU, Global Precipitation Climatology Centre (GPCC) and Global Precipitation Climatology Project (GPCP) in Africa, except in some regions where ground stations were very dispersed such as the Democratic Republic of the Congo and Angola (Harris et al., 2020). Middle correlation values (0.63) have been reported between CRU and ground stations concerning the inter-annual precipitation variability (Gbobaniyi et al., 2014).

Only temperature and precipitation have been analysed in this study (Figure 4.1). These two variables were used to calculate the EOF from which the present predictors of the crop yield were derived. Data from 1961 to 2019 for the region extending from 20° west to 55° east, and 14° south to 20° north were retrieved.

4.1.2 Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4)

The sea surface temperature is an important variable of climate fluctuation. It has been much used in climate modelling, assessment and monitoring. In this study, we used it as parts of the large-scale predictor to model the impact of climate change on maize yield. We analysed the SST anomaly extending from 15° east to 60° east and 20° south to 20° north in the Indian Ocean, and from 30° west to 15° east and 20° south to 20° north in the Atlantic Ocean (Figure 4.2).

Many SST products from a variety of institutions have been released (Parker et al., 1994; Kaplan et al., 1998; Reynolds et al., 2002; Ray et al., 2012), but in this study, we opted for ERSST.v4, which is the current, updated version of ERSST. The characteristics of ERSST.v4 are explained by Huang et al. (2015). In fact, the most important innovation compared to the previous version is the bias adjustment for night-time marine air temperature data. In general, ERSST are reconstructed from different dataset sources: observational data from the International Comprehensive Ocean–Atmosphere Data Set that covers the period from 1875 to 2007; the night marine air temperatures, covering the period from 1956 to 2010; the sea ice concentration from the Hadley Centre SST (HadISST) covering the period from 1870 to 2010; and the optimum interpolation SST dataset based on satellite observation and field measurements and covering the period from 1982 to 2011.

ERSST.v4 have been compared to direct observational datasets – such as the one produced by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)

– and to other reconstructed datasets. ERSST.v4 exhibited a warmer SST at about 0.2 $°C$ in the Southern Ocean, 0.1 $°C$ warmer in the Northern Ocean, and warmer in the lower latitude, especially in the eastern equatorial Pacific where the difference reaches 0.4 $°C$. The overall RSME of ERSST.v4 SST compared to ASTER is relatively low, estimated at $0.54°C$. It is important to mention that ASTER is the most accurate SST satellite estimate (Merchant et al., 2012). Compared to other reconstructed datasets, the SST of ERSST.v4 is similar to the previous version of ERSST and other datasets regarding the spatial mean, variability and trend. The spatial difference may be attributed to the new bias adjustment of ERSST.v4. For example, in the tropics, the SST anomalies are lower in ERSST.v4 than HadSST3. Compared to the previous version, ERSST.v3, ERSST.v4 exhibits a lower SST anomaly from 1920 to 1940. More details on the comparison of ERSST.v4 and other SST products are provided by Huang et al. (2015).

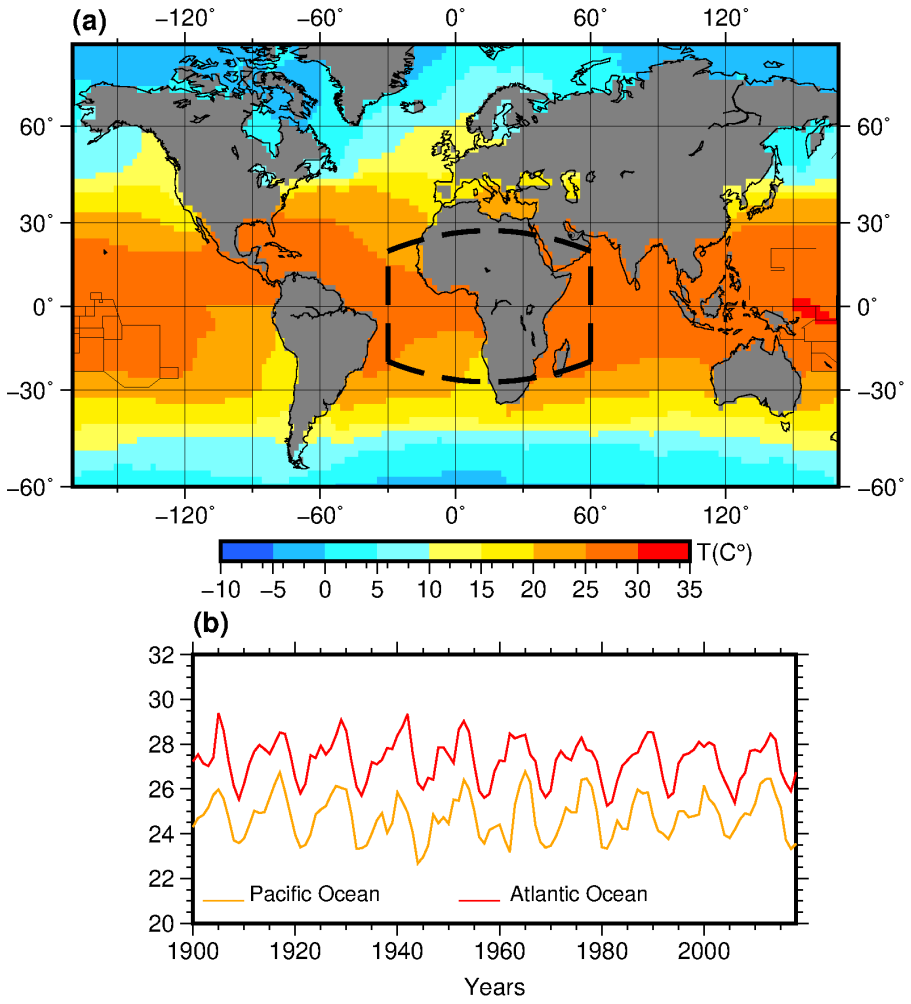

Figure 4.2: (a) Spatial climatology of mean SST; and (b) SST times series for the Pacific and Atlantic Oceans of the concerned region, defined by the dashed black line in the spatial pattern. The data come from ERSST.v4.

4.1.3 Reanalysis data (ERA5)

Reanalysis data sets are considered as an alternative to gridded observational data, especially in Central Africa where stations are sparse (Hua et al., 2019). They are preferred because they provide consistent and comprehensive datasets. Reanalysis produces the best estimate of the past atmospheric state by combining observational data and global forecasts within an assimilation scheme. Observational data come from different sources: land surface, ships, radiosondes, aircraft or satellites. The forecast model transfers the in-

formation from the previous analysis of the atmospheric state. The assimilation scheme combines short-range forecast with station observations to produce atmospheric state without spatial and temporal gaps, which will forecast the future. The reanalysis data can be obtained from different time steps (hourly, daily, monthly) and cover many climate variables. There are currently more than 10 available reanalysis datasets in the world (Hoffmann et al., 2019), including ERA5.

ERA5 is the fifth generation of the atmospheric reanalysis data produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). It covers the period from January 1979 to the present and will soon be extended back to 1950. The data are implemented through the integrated forecasting system cycle $41r2$ with four-dimensional variational analysis (Hoffmann et al., 2019). It has an horizontal resolution of ~ 31 km (TL639 spectral grid) and 137 hybrid sigma-pressure levels in the vertical, with the top level located at 0.01 hPa (an altitude of about 80 km). The data is available on a 0.25-*degree* grid with hourly intervals. Compared to previous versions of ECMWF, ERA5 outperforms ERA-Interim by improving the temporal and spatial resolution. Furthermore, the geophysical processes in the forecast model and the number of observation stations for the data assimilation system are improved. However, the drawbacks of ERA5 are that it requires more storage space and is computationally intense (Hoffmann et al., 2019). According to the recent result of Gleixner et al. (2020), ERA5 precipitation and temperature bias in Africa is negligible compared to other reanalysis data and the interannual variability over Africa is better represented. The spatial distribution is also better represented (Gleixner et al., 2020).

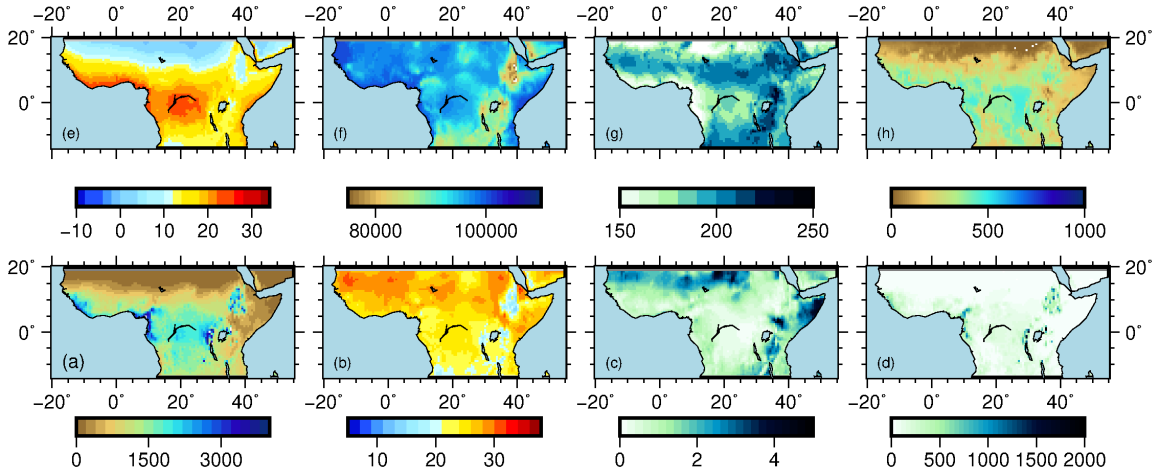

Figure 4.3: Climatology yearly means of selected variables: (a) total precipitation in mm; (b) mean temperature in $°C$; (c) 10-meter wind in m/s; (d) surface runoff in $mm/year$; (e) dew point temperature in $°C$; (f) surface pressure in Pa; (g) solar radiance $MJ.m.d^{-1}$; and (h) soil moisture in $mm/year$.

In this study, ERA5 climate variables have been used to calculate the WRSI, as demonstrated in chapter 6. These variables are: mean temperature, precipitation, maximum temperature, minimum temperature, 10 m wind, surface runoff, net solar radiance, surface pressure and soil moisture. We retrieved data from 1979 to 2019 with a region extending from 20° west to 55° east, and 14° south to 20° north (Figure 4.3).

4.2 Climate model data

Climate models were first designed to understand climate systems. They can also be used to study the future and changes in the past. According to Paeth et al. (2005), climate models provide many advantages over gridded observation datasets; they are not affected by measurement errors and provide continuous datasets. However, they are limited in capturing all the information related to the climate system because of the computational cost. Furthermore, they are very useful for understanding climate systems through numerous sensitivity analyses. Many climate models have been developed; some are assembled in the CMIP5, from which two were used in this study: ECHAM6 and REMO.

4.2.1 ECHAM6

ECHAM6 is the sixth generation of the coupled atmosphere ocean circulation model (Popke et al., 2013). The model was developed by the Max Planck Institute for Meteorology. It is the improved and most recent version of the previous ECHAM derived from the **E**uropean **C**enter for Medium-Range Weather Forecasts (the origin of **EC** in acronym). The model was initially developed in **Ham**burg (the origin of **HAM** in the acronym). With the CMIP5 experiment, ECHAM6 was coupled with the Max Planck Institute ocean model (MPIOM) (Jungclaus et al., 2013), the Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg (JSBACH) terrestrial biosphere model (Reick et al., 2013) and the Hamburg ocean carbon cycle model version 5 (HAMOCC5) (Ilyina et al., 2013).

The specificity of ECHAM6 is that it improves the large-scale circulation and the diabatic processes. The model is composed of a boundary dataset, a scheme of physical parametrisations of the diabatic processes, a spectral-transform dynamical core and a transport model. It uses the mixed finite difference/spectral discretisation of the primitive equation to address the adiabatic process. The vorticity and divergence form of the primitive equations drive the dynamical core of ECHAM6. The method of Simmons and Burridge (1981) is used for the vertical discretisation, and hybrid sigma-pressure is used on a Lorenz grid. The tracer transport is represented by the flux-form semi-Lagrangian scheme of Lin and Rood (1996).

The diabatic process of ECHAM6 is quite similar to the older versions, except for the land surface process and the treatment of solar radiative transfer. The eddy diffusivity/viscosity approach is applied to the turbulence parametrisation and boundary layer. The mass-flux framework developed by Tiedtke (1989) is applied to the parameterisation of moist cumulus convection. In contrast to previous versions, the condition for convection is based on the predicted variance of the virtual potential temperature, rather than constant temperature increments. The sub-grid orography scheme developed by Lott (1999) is used to parameterise the momentum transport. The Doppler spread theory is used to parameterise the unresolved vertical momentum transport disturbances. The humidity distribution function scheme implemented by Sundqvist et al. (1989) is used to describe the subgrid-scale cloudiness. ECHAM6 uses the most recent radiation transfer suite of models described by Iacono et al. (2008). The radiation transfer suite model is applied for both shortwave and longwave parts of the spectrum. This is the most significant difference from ECHAM5, which uses the Fouquart et al. (1980) model for the shortwave.

In ECHAM6, the rapid radiation transfer model of the shortwave scheme is based on the correlation k-method. The land surface model is composed of twelve types of plant functions and two bare soil types. The dynamical vegetation comes from the JSBACH terrestrial biosphere model where the land-use transition approach took the place of the land cover change.The model dissociates the visible albedo from the near-infrared one (Brovkin et al., 2009). In contrast to previous versions, ECHAM6 accounts for the solar zenith angle dependence over open water, for pond melt on the sea ice and for snow cover. A simple single-layer bucket model is used to present the soil hydrology, and the temperature is estimated over the five soil layers (Brovkin et al., 2009). A single value represents all the greenhouse gases, except water vapour and ozone. Carbon dioxide can also be computed in the Earth System model (ESEM), which is based on a biogeochemical component. ESEM describes the carbon fluxes between terrestrial, ocean and atmospheric reservoirs (Taylor et al., 2012). The time-evolving atmospheric concentration of carbon dioxide is therefore interactively calculated by the ESEM (Taylor et al., 2012). With the CMIP5 experiment, the model uses the photodissociation and oxidation of methane parameterisation to represent the water vapor in the middle atmosphere (Schmidt et al., 2006). The ozone is given as a three-dimensional monthly value, but it varies with the longitude in the stratosphere (Cionni et al., 2011).

According to Stevens et al. (2012), ECHAM6 yields a low standardised error compared to its predecessors. However, like other coupled atmosphere ocean models, ECHAM6 is not good at representing the sea surface temperature gyres in the upwelling regions west of the continents. Important biases were also reported in the Kuroshio Current region. A warm bias was observed in the upwelling region in the west of Africa. The strongest precipitation bias was reported in tropical Africa. As far as the tropical mode variability is concerned, ECHAM6 and its predecessors simulate the Madden-Julian-Oscillation (MJO) quite well (Schubert et al., 2013). It outperforms its predecessors in representing the ENSO, and the amplitude is much improved. Other characteristics of the ENSO, such as the principal spatial pattern, are also well represented, except in the west, where the sea surface temperature variability pattern stretches out. The Atlantic Ocean teleconnection is also well simulated. However, the model poorly represents the central subtropical North and South Pacific and tropical western Pacific teleconnections.

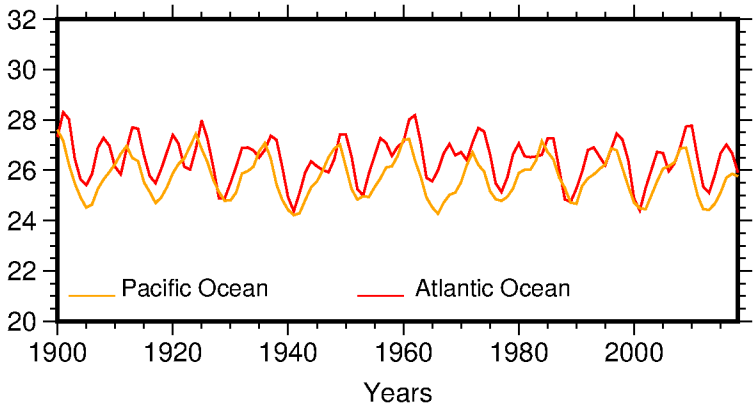

Figure 4.4: ECHAM6 SST from 1900 to 2020

The model has been run with a high emission scenario (RCP 8.5) and a midrange mitigation emissions scenario (RCP 4.5). We used the low-resolution configuration, with similar horizontal resolution to ECHAM5/MPIOM. The vertical extends to 0.01 hPa, capable of representing the variability of the high latitude circulation in the middle atmosphere. We are only interested in SST from 1979 to 2100 (Figure 4.4).

4.2.2 REMO

REMO is a limited-area three-dimensional atmospheric circulation model based on the 'Europa-Modell' of the German weather service. The Europa-Modell constitutes the dynamical core of the REMO model, where the primitive equations are resolved. The second core is based on the parameterisation scheme inherited from ECHAM, in our case, ECHAM6, where important physical processes of the atmosphere are not resolved in the dynamic equations.

The prognostic variables involved in the primitive equation are horizontal wind component, water vapour component, surface pressure, cloud water and temperature. The model uses the lateral boundary suggested by Davies (1976), which involves the relaxation of the interior flow in the vicinity of the boundary to the external fully prescribed flow. It is based on the five-layer soil model (down to 10 m depth) to simulate the land surface. The vegetation parameters are derived from the new version of the land surface parameters, known as the LSP2 dataset. This new version improves the background surface albedo of the bare soil of Africa (Hagemann, 2002). In REMO, the fourth-order

linear horizontal diffusion scheme is applied to momentum, temperature and water content. The subgrid-scale atmospheric processes, such as radiation, convection or clouds, are not resolved and must be parameterised. The micro-cloud scheme has been described by Lohmann et al. (1996). This new approach of the micro-cloud scheme dissociates the water cloud prognostic variable to the cloud ice and the maritime to continental clouds. The model uses turbulent kinetic energy to calculate the eddy diffusion coefficients. The vertical diffusion and turbulent surface fluxes are calculated with a higher-order closure scheme for the transfer coefficients of momentum, heat, moisture and cloud water within and above the planetary boundary layer from the Monin–Obukhov similarity theory (Louis, 1979). The radiation scheme is based on the narrow band models of longwave radiation transfer, as described by Morcrette et al. (1986), but with a slight improvement for additional GHG, 14.6 μm band of ozone and various types of aerosols. The moisture convection is derived from the mass flux scheme, as explained by Tiedtke (1989).

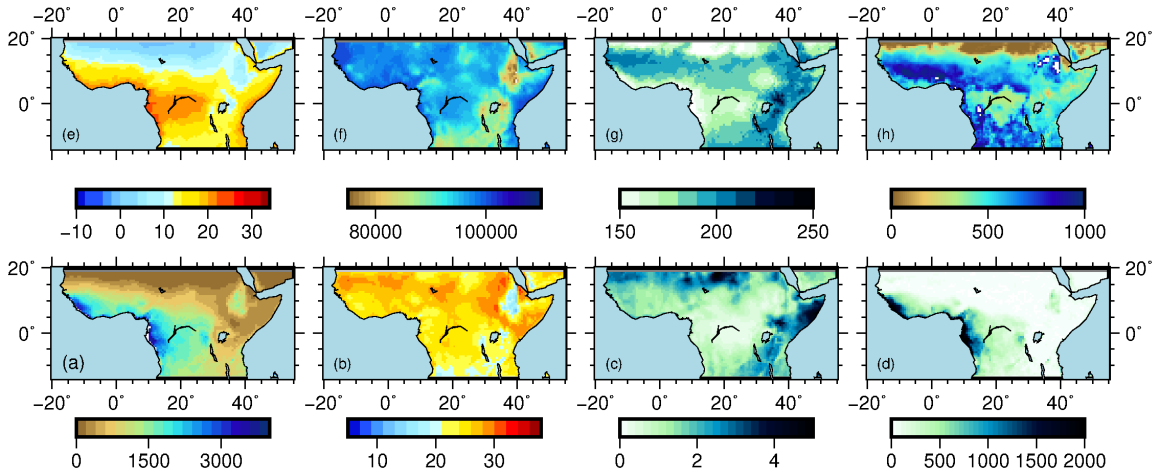

Figure 4.5: Spatial climatology of selected variables from historical REMO (1980-2005): (a) total precipitation in mm; (b) mean temperature in $°C$; (c) 10-meter wind in m/s; (d) surface runoff in $mm/year$; (e) dew point temperature in $°C$; (f) surface pressure in Pa; (g) solar radiance $MJ.m.d^{-1}$; and (h) soil moisture in $mm/year$.

As far as numerical calculation of the dynamic part is concerned, the temporal discretisation is implemented by a leap-frog scheme with semi-implicit correction. The vertical is discretised by hybrid σ -coordinates, as described by Simmons and Burridge (1981). In the horizontal plane, the spherical Arakawa-C grid is used where prognostic variables are adjusted in a boundary zone of eight grid boxes and calculated separately for the different compartments within a model grid box. The model is run in a rotated grid to

have similar grid box sizes and an optimised time step.

As for other regional climate models, REMO can be driven by different boundary conditions, including perfect boundary conditions from ECMWF 40-year reanalysis. In this study, we opted for the REMO MPI-CSC version, developed at the Max-Planck Institute for Meteorology and the German Climate Service Centre in 2009 and was part of CMIP5. With this version, REMO is nested into the radiative transfer scheme of the atmospheric component of MPI-ESM-LR (atmospheric general circulation model ECHAM6) developed at the Max Planck Institute (MPI) in Hamburg (Stevens et al., 2012). The 0.44 grid degree resolution (almost 50 km horizontal resolution) was interpolated into a 0.5-degree resolution, similar to observational data. The entire model domain of REMO extends from 30° west to 60° east and from 15° south to 45° north. As for observational datasets used in this study, only datasets from the selected region, extending from 20° east to 55° east, and from 14° south to 20° north, were analysed. The historical experiment of REMO in this study covers the period from 1979 to 2005 and the projection covers the period from 2006 to 2100.

The result of the REMO simulation or any other RCM is validated by observational datasets. Many researchers have reported good performance of REMO in simulating major features over tropical Africa. According to Paeth et al. (2005), REMO successfully simulates the main features of observed mean, trend and inter-annual variability, as well as large-scale circulation over West Africa. However, the model cannot effectively reproduce the sub-Saharan precipitation because of the coarse resolution of the Guinean Coast orography (Paeth et al., 2005). More recently, Fotso-Nguemo et al. (2017a) found that REMO could better represent the spatial distribution of precipitation and temperature over Central Africa. They found correlation coefficients of 0.74 for precipitation and 0.93 for temperature between REMO and CRU datasets. In the same direction, Fotso-Nguemo et al. (2017b) found that REMO was able to simulate quite well the precipitation inter-annual variability, the daily precipitation intensity distribution and the intra-seasonal variability of the Central African monsoon. Vondou and Haensler (2017) reported the ability of the model to represent the variability in precipitation anomalies between different events associated with El Niño in Central Africa. Finally, Tamoffo et al. (2019b) found the model reasonably simulated the frequency of wet days, the threshold of extreme precipitation and the cumulative frequency of daily precipitation over Central Africa.

4.3 Climate teleconnection patterns and indices

Teleconnection patterns refer to a strong negative correlation between two or more centers of action, mostly between high and low-pressure regions. These opposite centers are separated by thousands of kilometers. The action centres present a monthly, yearly or decennially synchronised behaviour pattern, forming a dipole, which comprises a region of high and low pressures in a specific region of the planet (Barnston and Livezey, 1987; Martinez-Artigas et al., 2021). This study uses the time series indices of climate teleconnection patterns to select and give the physical interpretation of the crop yield predictors derived from the EOF analysis. They have been chosen according to their role in the variability in the tropical African climate. Thus, we targeted the ENSO, the North Atlantic oscillation (NAO), the IOD, the South Atlantic Ocean dipole (SAOD), the Indian summer monsoon (ISM), the Southern Annular Mode (SAM), the tropical North Atlantic index (TNA) and the tropical South Atlantic index (TSA). Teleconnection patterns were obtained from the NOAA Physical Sciences Laboratory (https://psl.noaa.gov/data/climateindices/).

ENSO is considered the most important mode that drives the inter-annual climate variability (Lindesay, 1988). ENSO has an impact on a global scale and determines the seasonal climate variability in the world. The influence of ENSO – which is happening in the central and eastern equatorial Pacific – can be observed in the polar regions of both hemispheres (Yuan et al., 2018). The teleconnection pattern can influence local climate variables such as precipitation and temperature. It has two phases: El Niño and La Niña. El Niño refers to the negative phase of the ENSO index, accompanied with ocean warnings on the western coast of South America. Usually, water on the western coast of South America is cold because of the Humboldt sea current that comes from Antarctic latitudes. The association of ENSO and the trade wind brings abundant precipitation and floods on the west coast of South America. In contrast, drought is reported in the opposite centre, Australia and Indonesia. Therefore the effect propagates to the rest of the world through Rossby waves, the jet stream, meridional and zonal circulation and eddy activities (Martinez-Artigas et al., 2021). Numerous studies have indicated a close correlation between local climate variables and teleconnection patterns. For example, in the east of Africa, a strong correlation between the ENSO and precipitation have been reported (Hastenrath and Greischar, 1993; Nicholson, 2017). ENSO has a high correlation with the short rain season (Indeje et al., 2000; Wolff et al., 2011; Bahaga et al., 2019) in East Africa. According to Rouault and Richard (2005), more than half of severe droughts in Southern Africa, for example, are attributed to ENSO. But the relationship between

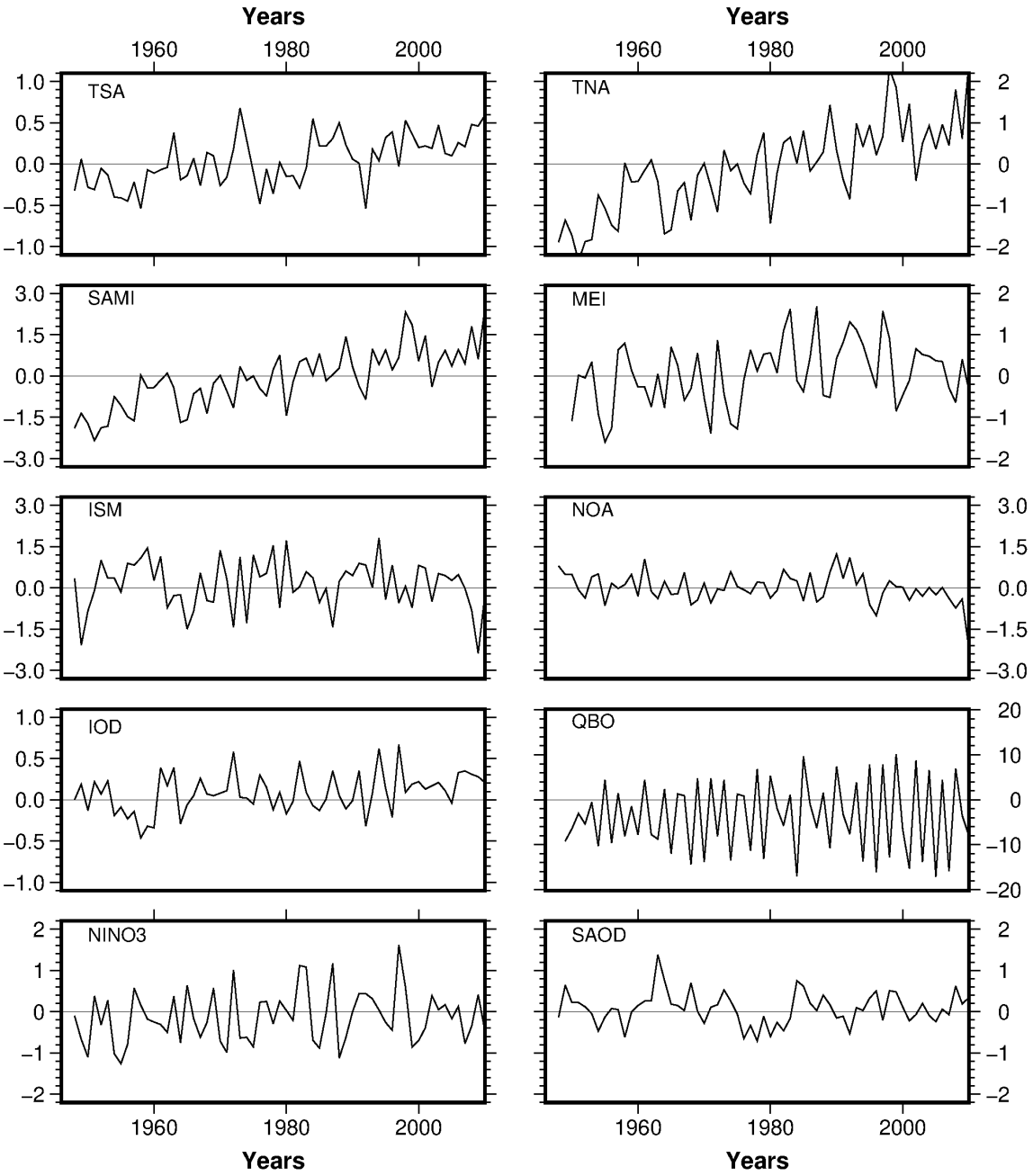

Figure 4.6: Climate indices and teleconnection patterns from 1948 to 2010. Source: https://psl.noaa.gov/data/climateindices/.

ENSO and local climate variables is not always linear (Reason and Jagadheesha, 2005; Yuan et al., 2014). For example, one of the strongest El Niño events in 1997 did not drastically impact Southern Africa's precipitation compared to moderate events (Wolff et al., 2011). In the west of Africa, Srivastava et al. (2019a) demonstrated that ENSO has a significant impact on the West African summer monsoon by regulating the temperature of the upper troposphere of subtropical Africa. Many other researchers found that ENSO could explain around 24 % of West African monsoon inter-annual variability (Losada et al., 2012; Mohino et al., 2011). In Central Africa, in contrast, the influence of ENSO is limited. Several indices have been suggested to define ENSO; in this study, we used Niño 3 and the multivariate ENSO index (MEI). Niño 3 is calculated in the region extending from 5 ° north to 5 ° south and from 150 ° west to 90 ° west. The MEI is calculated based on five variables: sea-level pressure, SST, zonal and meridional components of the surface wind and outgoing longwave radiation (Wolter, 1987).

The Quasi-biannual Oscillation (QBO) is a quasi-periodic movement of the equatorial zonal wind from east to west, with a period of 28 or 29 months. This phenomenon happens at the top of the low tropical stratosphere and converges to the tropopause with a velocity of 1 km per month, where it will dissipate. There are not many studies investigating the association of QBO and climate variables. We incorporated QBO in this study simply because Mason and Tyson (1992) found a strong correlation between QBO and Southern African precipitation. Furthermore, Ng'ongolo et al. (2010) has improved the statistical perdition of East African precipitation using quasi-biennial oscillation.

NAO is one of the most dominant teleconnection patterns that drives the climate variability over the North Atlantic Ocean. Its impact on temperature and precipitation is greatly pronounced in north eastern America, the North Atlantic and the Eurasian continent. The positive or neutral phase of NAO indicates that low pressure dominates the high latitudes in the North Atlantic. In contrast, high pressure dominates the opposite action centre located between the Mediterranean region and North Africa. In Africa, the effect of NAO on the inter-annual precipitation variability has been observed in Morocco (Lamb and Peppler, 1987). Moreover, (McGregor, 2015) analysed 31 meteorological stations, extending from 0° to 16° south and 25° to 40° east and found significant negative correlation between all stations and the NAO index.

IOD is the predominant teleconnection pattern on the inter-annual time scale in the Indian Ocean. Its positive phase is characterised by a cooling and warming water in the eastern and western equatorial Indian Ocean respectively. Enhanced convection is therefore observed in the west, resulting in abundant precipitation in East Africa, mostly

during the short rain season. Its impact has been much documented in East Africa, between 10° south and 10° north and extends from the Indian east coast up to 25° east (Gadgil et al., 2004). According to the recent study of (Srivastava et al., 2019c), IOD could explain about 80 % of precipitation during the short rain season inter-annual variability in East Africa.

SAOD is a teleconnection pattern that influences the sea surface temperature variability in the Atlantic Ocean. SAOD has two centres of action: one at the Argentina–Uruguay–Brazilian coast; another at the central equatorial and West African coast, also referred to as the Atlantic Niño region. Its impact on precipitation has been assessed in West Africa through numerical model simulations (Kucharski et al., 2009; Losada et al., 2010). Nnamchi et al. (2011) demonstrated, using observational climate datasets, that the positive phase of SAOD is associated with the occurrence of positive precipitation anomalies over 40 mm per month in the Guinea Coast region.

ISM is a coupled land–ocean–atmosphere phenomenon that plays an important role in bringing moisture to the Indian subcontinent (Naidu et al., 2020). The phenomenon does not affect the climate of Africa, except in some Mediterranean regions. According to Rizou et al. (2018), the effect of ISM in Mediterranean regions can be attributed to the propagation of the equatorial Rossby wave. This does not concern tropical Africa; however, we used the ISM to control the robustness of predictors not to capture climate phenomena happening elsewhere.

SAM, sometimes referred to as the Antarctic Oscillation, is the leading teleconnection pattern that drives climate variability over the Southern Hemisphere region (Fogt and Marshall, 2020). The index corresponds to the first principal component of mean sea level pressure from gridded datasets. It explains between 22 and 32 % of the extratropic Southern Hemisphere climate variability, according to the observational dataset used (Fogt and Marshall, 2020). In Africa, SAM index has only has influence on the Southern African climate (Reason and Jagadheesha, 2005).

Two more additional climate indices have been used in this study: the TNA and the TSA. TNA refers to the anomaly of the average of the monthly SST from 5.5 ° north to 23.5 ° north and 15° west to 57.5° west, while TSA are calculated from the equatorial line to 20 °south and 10 °east-30° west. These two indices have an impact on the variability of the tropical African climate. Worou et al. (2020) conducted a systematic study of all possible climate indices that could have an impact on precipitation in West Africa and found a strong correlation coefficient (0.58) between TSA and the Guinean Coast precipitation

index.

4.4 Crop yield dataset

Crop yield datasets generally come from national statistics and international agencies or research institutions. Depending on the country's organisation, national crop yield data are usually collected per crop at village level in local agricultural centres. These datasets are further aggerated at the district, provincial and national levels. International organisations such as the FAO refer to these datasets to construct a continuous and standardised database. However, national datasets are most of the time incomplete, and sometimes inconstant. In this respect, the FAO uses the imputation technique to improve national datasets. This technique leads to the reduction of the internal variability of the dataset (as illustrated in Figure 4.7 for maize in Democratic Republic of Congo) and makes climate variability studies on crop yields difficult. With the development of satellite devices, gridded datasets are becoming popular and are used to calibrate and simulate crop yield models. They merge crop statistics datasets with existing biophysical parameters. Two important gridded crop yields datasets are commonly used in the scientific community: the one developed by Iizumi et al. (2014), referred to here as the GDHY dataset, and the one compiled by Ray et al. (2012).

This study uses the most recent version of the GDHY dataset because it is consistent (Figure 4.8). This version is referred to as GDHY v1.2 + 1.3. It was released in 2020 and covers the period from 1981 to 2016, with a 0.5° spatial resolution. GDHY v1.2 + 1.3 is an aligned version of two datasets: GDHY v1.2 and GDHY v1.3, which cover the periods from 1981 to 2011 and from 2000 to 2016, respectively. All three versions used the census statistics from the FAOSTAT as primary input and have as additional inputs:

- The harvested area derived from M3-Crops (0.083° and average around 2000);
- the crop calendar data, implemented by Sacks et al. (2010);
- the production share dataset developed by the USDA;
- and the satellite products.

Many global crop yield studies are based on the GDHY v1.2 dataset (Iizumi et al., 2014; Iizumi and Ramankutty, 2016). Apart from the difference in coverage period, the

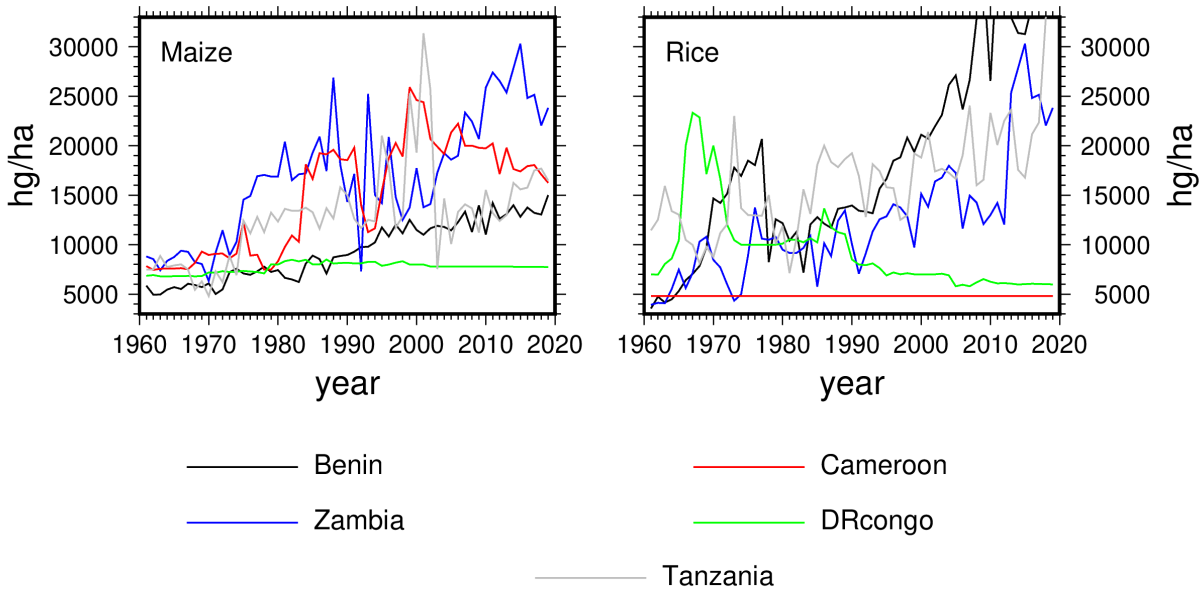

Figure 4.7: Crop yield times series data from 1960 to 2020 for selected African countries. Data source: FAOSTAT.

GDHY v1.2 differs from the GDHY v1.3 by the type of satellite products and solar radiation datasets. GDHY v1.2 uses the vegetation leaf area index and a fraction of photosynthetically active radiation derived from global inventory modelling and mapping studies and the solar radiation obtained from the JRA-25 reanalysis dataset to calculate the grid cell net primary production, which was associated with FAOSTAT census data (Iizumi and Sakai, 2020). In contrast, the grid cell net primary production of the GDHY v1.3 dataset is derived from the moderate resolution imaging spectro-radiometer and JRA-55 reanalysis for solar radiation.

The alignment of both datasets was necessary to create a continuous and longer dataset, which meets the expectation of the scientific community, agribusiness, crop modellers and food agencies, especially for times series analysis. Iizumi and Sakai (2020) have described in detail this alignment method. The alignment results were validated by an independent observed census dataset assembled by Ray et al. (2012). The result indicated that the GDHY yielded a good agreement with the independent census.

It is important to mention that the accuracy of all GDHY datasets depends on regions, especially where FAOSTAT census data are missing. Also, the GDHY dataset is not

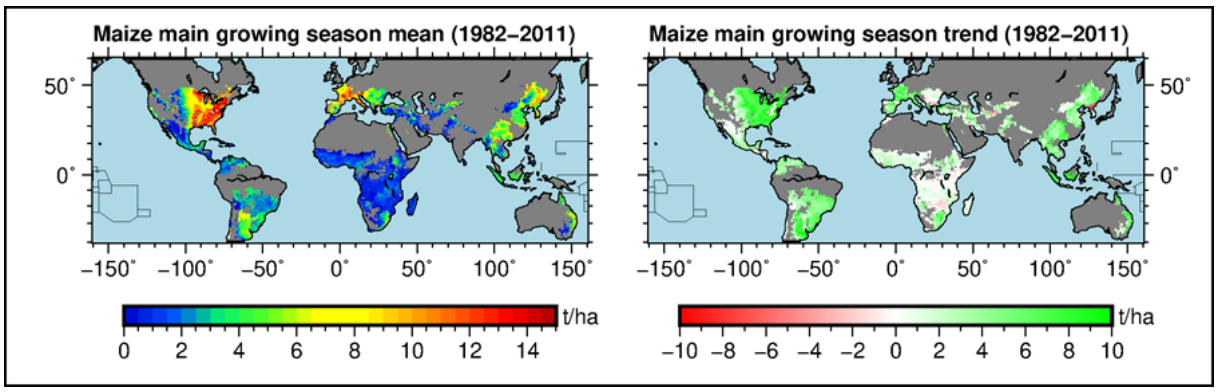

Figure 4.8: Gridded crop yield data: annual mean (left) and linear trend (right) from 1982 to 2011.

purely considered as an observed dataset, but an estimated one, therefore also subject to biases. Uncertainties related to GDHY v1.2 global yield gaps, trend patterns, growth rates and changes in year-to-year variability are clearly described by Iizumi and Ramankutty (2016). Despite these imperfections, the current GDHY data and its previous version are the most widely used in scientific communities, especially in climate change impact studies on crop yield, and have served as calibration and validation datasets of many crop simulation models. Iizumi and Sakai (2020) studied the impact of climate change on yield variability using the GDHY v1.2 dataset. Iizumi et al. (2014) assessed the impact of ENSO on global crop yield using the same dataset.

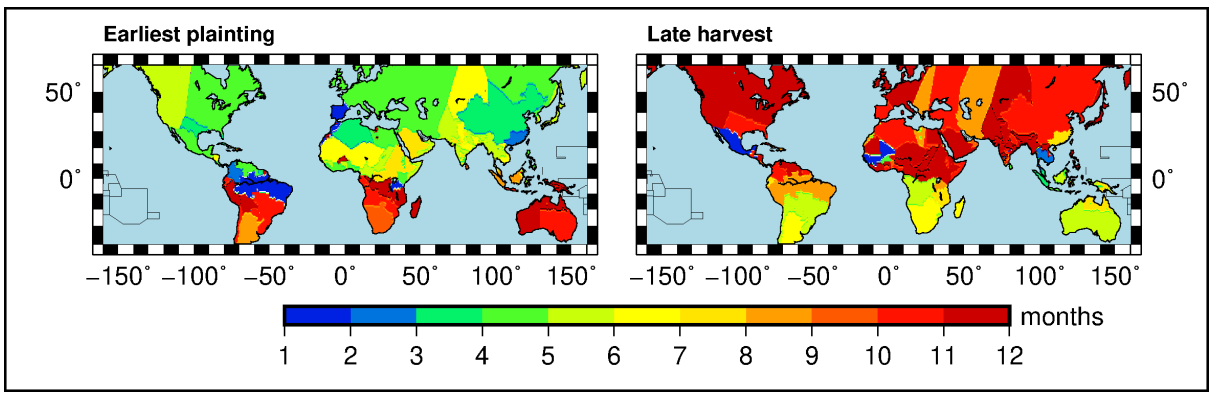

Figure 4.9: Cropping calendar for maize yield.

In this study, we also used the agricultural crop calendar implemented by Sacks et al. (2010) to restrict the model to growing seasons only (Figure 4.9). The data were derived from observational crop planting and harvesting dates, mainly collected by the

United Nations Food and Agriculture Organization and the United States Department of Agriculture. It contains more than 1300 observational planting and harvesting days for 19 crops; in our study, we are only concerned with two crops: maize and rice Sacks et al. (2010). The crop calendars were obtained in graphical formats, then digitised and geo-referenced. In tropical Africa, the general pattern of planting dates for many crops depends on precipitation and the rapport between precipitation and evapotranspiration, while in precipitation-limited regions such as Kenya, the onset of the rainy season determines the planting date. In some regions, the planting data is set based on the flowing periods Sacks et al. (2010). In Nigeria for example, maize planting occurs in June to allow the flowing period to match the rainy season. Other factors not related to climate can also determine the plating date, such as tractor ownership Sacks et al. (2010). For maize, both the major and secondary growing seasons are of concern. This study deals only with main growing season.

5. Statistical methods and data pre-processing

This chapter describes in detail the different statistical methods used to obtain the thesis results. It also addresses some data pre-processing techniques.

5.1 Data pre-processing

The pre-processing of raw data is necessary before developing the crop yield model. The following section presents two techniques used to prepare the dataset: interpolation and statistical bias correction.

5.1.1 Spatial Interpolation

As described in section 4, the land climate dataset and the maize yield dataset came from different data sources with different spatial resolution and different position of the grid point centre. REMO has a spatial resolution approximating 0.44°. Both CRU datasets and crop yield have a spatial resolution of 0.5°. The reanalysis ERA5, from which local variables have been taken, comes with a finer spatial resolution of 0.1° . Because the model is run in each grid point, both ERA5 and REMO are interpolated to the crop yield grid. Several algorithms have been suggested to conduct such an operation, but the inverse distance weighting interpolation was applied in this study. The inverse distance weighting is a deterministic interpolation that uses an advanced nearest neighbour algorithm to predict the unknown value by weighting the nearest neighbours surrounding the given point (Hartkamp et al., 1999). It is widely used in environmental science and climatology (Samanta et al., 2012). In this study, the interpolation was based on six neighbours. It

is therefore mathematically computed as the following:

$$Y(t, s_y) = \frac{1}{\sum_{s_x=1}^{n} w(s_x)} \sum_{s_x=1}^{n} w(s_x).X(t, s_x) \tag{5.1}$$

Where $Y(t, s_y)$ represents the interpolated variable in the new grid space and $X(t, s_x)$ a variable in the original grid space, for each new location s_y at the time t, s_x represents the known values at a nearby location. $w(s_x)$ is the weight function; it is an inverse distance power function of the linear distance between the value points. It decreases when the distance increases and is estimated as the following:

$$w(s_x) = (\frac{1}{d(s_y,\ s_x)})^{d} \tag{5.2}$$

Where d is the power parameter of the interpolation. It is generally set to two, as the case for this study.

5.1.2 Quantile mapping bias correction

Bias correction is an important step for climate impact studies, especially when climate variables are used directly in the regression model as the case for local predictors in this study. The bias correction follows three steps: the application of the transfer function to the simulated dataset for the control period; the application of the transfer function to the future simulated datasets; and the comparison between the corrected data and the observation. We used the quantile mapping bias correction.

In general, quantile mapping methods implement statistical transformations for post-processing of climate modelling outputs (Enayati et al., 2021). The statistical transformations involve transforming the distribution functions of the modelled variables into the observed ones. This method corrects the shape and the variability of the simulated model and is based on the empirical transformation of Panofsky and Brier (1968). It was initially applied in hydrology and has become the most widely applied bias correction technique. According to Piani et al. (2010), the statistical function of the quantile mapping can be mathematically expressed as follows:

$$x^0 = f(x^m) \tag{5.3}$$

In which x^0 is the observed variable, x^m is the modelled variable, and $f()$ is the transformation function. The objective here is to find a function f that maps the modelled variable x^m so that its new distribution equals the observed variable distribution x^0. In the context of this study, the observed variables are the local climate variables obtained from ERA5, and the modelled are those that come from REMO. The cumulative density function of the observed data and the simulated data can be obtained according to this mathematical expression:

$$x^0 = F_0^{-1}[F_m(x^m)] \tag{5.4}$$

Where $F_m(x^m)]$ represents the cumulative density function of x^m, and $F_0^{-1}[]$ is the inverse form of the cumulative density function of x^0, which is technically assimilated to the quantile function.

There are many variants of quantile mapping; Galmarini et al. (2019) enumerated that they are mostly applied in crop yield modelling. Depending on the target climate variable, the quantile mapping can use either a parametric or a non-parametric approach. A parametric approach uses the parametric transformation functions to form the quantile-quantile relation. It is therefore applied when the theoretical distribution of the variable is known, for example, in the normal distribution of the temperature and the gamma distribution for the precipitation (Themeßl et al., 2012). Different variations of the parametric transformation of the quantile-quantile relation can be mathematically expressed as such:

$$P_0 = bP_m \tag{5.5}$$

$$P_0 = a + bP_m \tag{5.6}$$

$$P_0 = bP_m{}^c \tag{5.7}$$

$$P_0 = b(P_m - y)^c \tag{5.8}$$

$$P_0 = (a + bP_m)(1 - e^{(-(P_m - y)/\varepsilon)}) \tag{5.9}$$

The P_0 represents the best estimate of P_m, and a, b, c, d, y and ε represent calibrated free parameters. By contrast, the empirical quantiles employ a non-parametric transformation function. The empirical cumulative distributions are estimated for the observed and simulated time series at fixed intervals. The empirical quantile mapping technique solves the cumulative density equation by using the cumulative density function of observed and modelled values instead of the suggested theoretical distributions. The empirical

cumulative density distribution is approximated using tables of empirical percentiles. The values between percentiles are estimated using linear interpolation. This process can be mathematically represented as such:

$$P^{Obs} = h(P^{RCMCon}) = ECDF^{Obs^{-1}}(ECDF^{RCMCon}(P^{RCMCon})) \tag{5.10}$$

The probabilities in $ECDF^{Obs}$ and $ECDF^{RCMCon}$ are approximated at a fixed interval. The h is calculated as the relative difference between the two $ECDFs$ in each interval.

After the calibration of the historical period, the transfer functions are then applied to the future climate change simulations – from 2006 to 2100 in the context of this study – under the assumption that they are time-invariant (Cannon et al., 2015). When values of the future climate projections are larger than the training values used to estimate the empirical cumulative density function, we used the correction found for the highest quantile of the training period (Themeßl et al., 2012). The statistical evaluation of the bias correction performance was not under the scope of this study. However, we computed the difference between the simulated and observed dataset to show areas prone to bias.

Figure 5.1 indicates how well the quantile mapping bias correction has improved the empirical cumulative density function of the temperature and precipitation of the simulated REMO under the RCP4.5 climate emission scenario. Comparing the observed temperature with the historical REMO (black line), between $13°C$ and $26°C$, the historical temperature of the REMO underestimates the observed temperature (green line) at about $1°C$. However, above $26°C$, they both look completely similar. The empirical quantile mapping bias correction has substantially corrected the simulated temperature (red line) in such a way that its cumulative density function coincides with that of the observed temperature. The benefit of the bias adjustment propagates to the future climate projection of the REMO under the RCP4.5 (orange line), which is warmer in all quantiles than the historical REMO. The bias correction has also improved the distribution of the simulated REMO precipitation (red line), where the uncorrected historical REMO underestimates the observed precipitation (green line) at about 1 mm/day between 1 and 7 mm/day. Above $7mm/day$, the REMO slightly overestimates the observed precipitation. As for the precipitation, the quantile mapping bias correction works well because the cumulative density function of the historical corrected REMO coincides with one of the observed datasets.

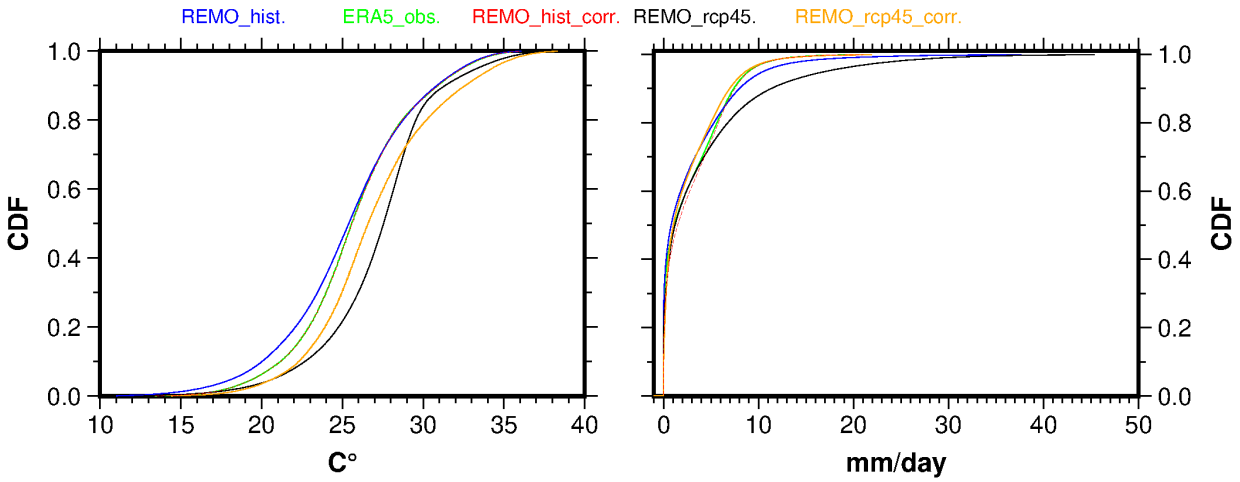

Figure 5.1: Cumulative density probability Function of the observed ERA5, corrected and uncorrected REMO. (Left) temperature, (right) precipitation.

Figure 5.2 compares the seasonal cycle of the observed ERA5 dataset to corrected and uncorrected REMO datasets. The uncorrected REMO (blue lines) exhibits a positive bias for surface runoff, soil moisture, 10 m wind and precipitation over the whole season. In contrast, a negative bias is observed for solar radiation and surface pressure for the entire seasonal cycle. The REMO temperature shows, on average, a positive bias of about 0.5 $°C$ compared to the observed data, but negative biases are observed in December, January and February. After correction, the positive bias substantially increased, but the negative bias increases and propagates to other months (March, April and May). The REMO precipitation bias indicates a wetter bias of 43 $mm/month$ and becomes 3.4 $mm/month$ after bias correction, but January, February and April display a negative bias. The REMO dew point temperature has, on average, a negative bias (-0.62 $°C$), but from July to October, REMO displays a positive bias. The average bias becomes -0.001$°C$ after correction. The REMO 10 m wind shows a positive bias (0.33 m/s on average). The bias completely disappears after the correction. Bias correction reduces the negative solar radiation bias from -14.3 to -1.2 J/day and surface pressure from -139 Pa for all months to -1.3 Pa. Stronger positive bias is observed for surface runoff (7.81 $mm/month$) and soil moisture (165 mm). After correction, the bias for surface runoff and soil moisture is reduced to -0.51 $mm/month$ and -0.33 mm.

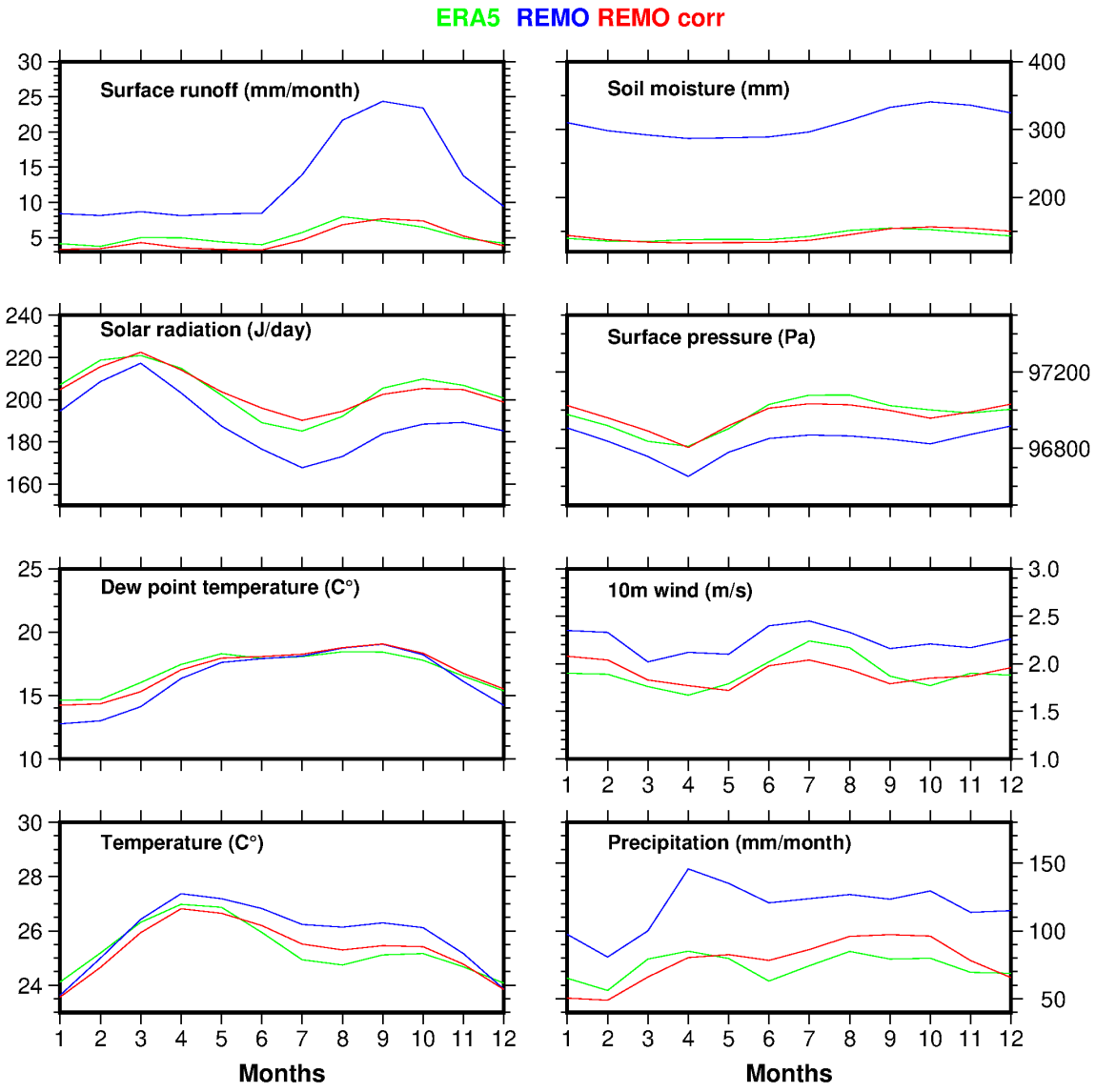

Figure 5.2: Annual cycle of area averaged-temperature, precipitation, 10 m wind, dew point temperature, solar radiation, surface pressure, soil moisture and surface runoff. The green curve corresponds to the observed dataset (ERA5), blue is the uncorrected historical REMO, and the red is the corrected historical REMO (with empirical quantile mapping bias correction approach.

Figure 5.3 displays the spatial pattern of ERA 5 data – simulated historical and corrected REMO. The corresponding biases are represented in Figure 5.4. According to the three datasets, abundant precipitation is observed in the Guinea coast, the Atlantic Ocean and the Congo Basin. The observational dataset, ER5 shows abundant precipitation in Ethiopia's highlands in the east of the Congo Basin and the mountains of Cameroon. In contrast REMO seems to underestimate the precipitation over these regions. The temperature distribution of REMO and ERA5 is quite similar. Regions with high temperatures are in the Sahel. The Ethiopian highland and rift valley are considered cold areas. The Somalian coast and the Sahel region are affected by strong 10 m winds around 4 m/s, as indicated by REMO and ERA 5. The spatial distribution of the surface runoff in REMO and ERA5 is different. REMO overestimates the surface runoff in the Atlantic coast and Guinea coast and underestimates it in the Ethiopian highlands and other regions. REMO totally overestimated soil moisture, except in the Congo Basin. The spatial distribution of REMO and ERA5 in representing solar radiance is similar. High values of solar radiance are observed in the east of Africa and in Sahel. Both datasets'surface pressure The surface pressure is similar, where high values are located at high altitude. The spatial distribution of dew point temperature in REMO and ERA5 is also similar. High values of dew point temperature are located in the Congo Basin.

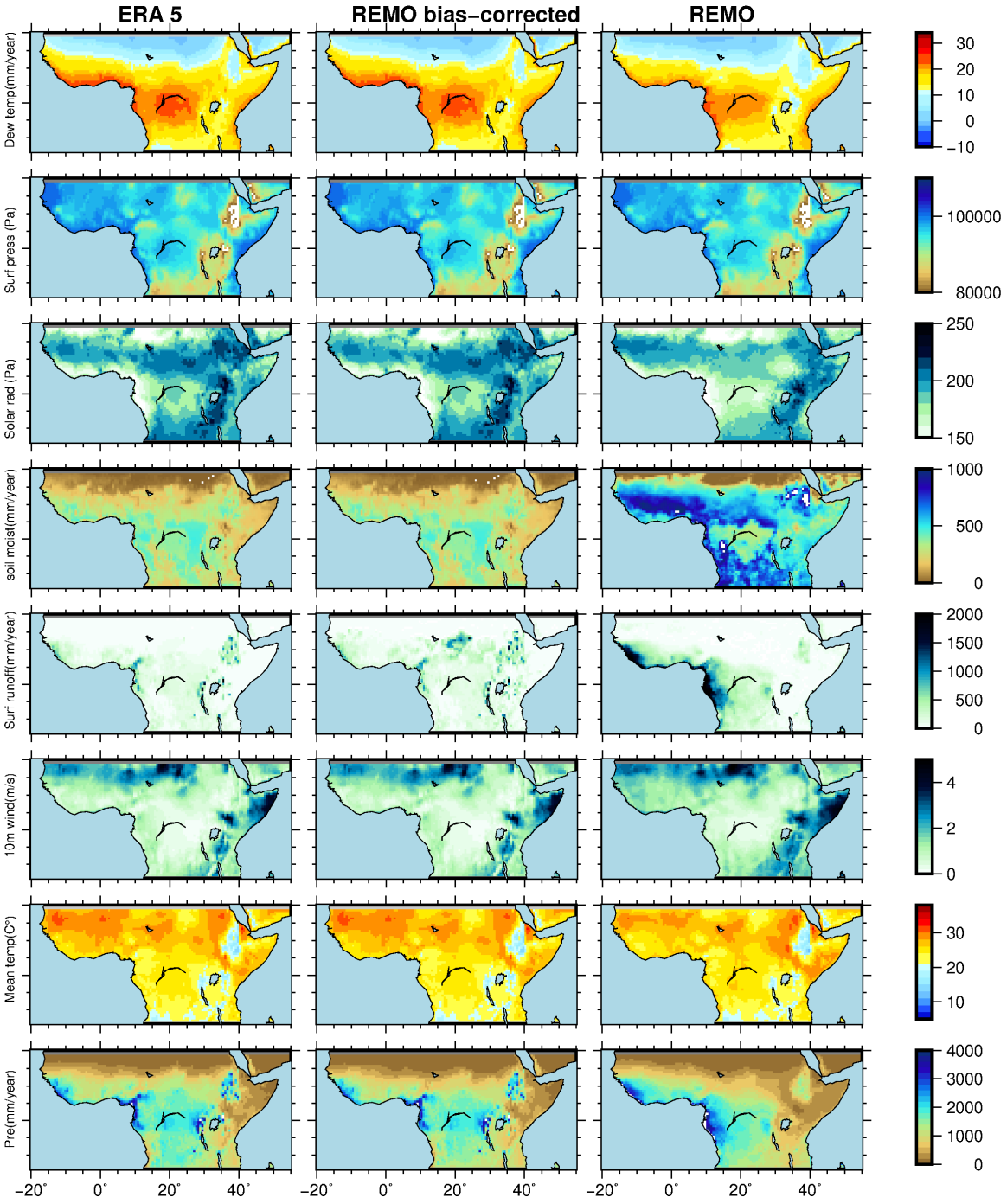

Figure 5.3: Spatial distribution of observation ERA5 (extreme left), historical raw simulated REMO (extreme right), and historical corrected REMO (middle) for mean temperature, mean precipitation, 10 m wind, dew point temperature, solar radiation, surface pressure, soil moisture and surface runoff.

Figure 5.4 displays the arithmetic difference between the historical REMO and the ERA5 (REMO minus ERA5). It completes Figure 5.3 by bringing detailed information, which was masked out; here, the spatial bias is quantified. A positive precipitation bias of around 500 $mm/year$ is observed in West Africa. In the west of the Congo Basin, a positive precipitation bias of more than 1000 $m/year$ is observed along the Atlantic coast. Meanwhile, the east is characterised by a negative bias, reaching 1200 $mm/year$ in the Ethiopian highlands and east African rift valleys. The surface runoff displays a similar bias pattern as the precipitation. The difference is that the negative bias in east Africa is low. The REMO temperature has a negative bias in the Sahel and West Africa, at around 2 $°C$. The 10 m wind displays a positive bias of 1 m/s almost everywhere, except on the west coast of the Congo Basin. This is similar to soil moisture, which displays a positive bias of 500 $mm/year$ everywhere, except in the centre of the Congo Basin. REMO displays a negative bias of 4 $°C$ regarding the dew point temperature in east Africa. In West Africa and in the Sahel, an average negative bias of 3 $°C$ is observed. REMO presents a strong negative bias of around 40 $Joules$ regarding solar radiance in east Africa in the west of the Ethiopian highlands. Meanwhile, positive biases are observed in the cost of Somalia (around 20 $Joules$).

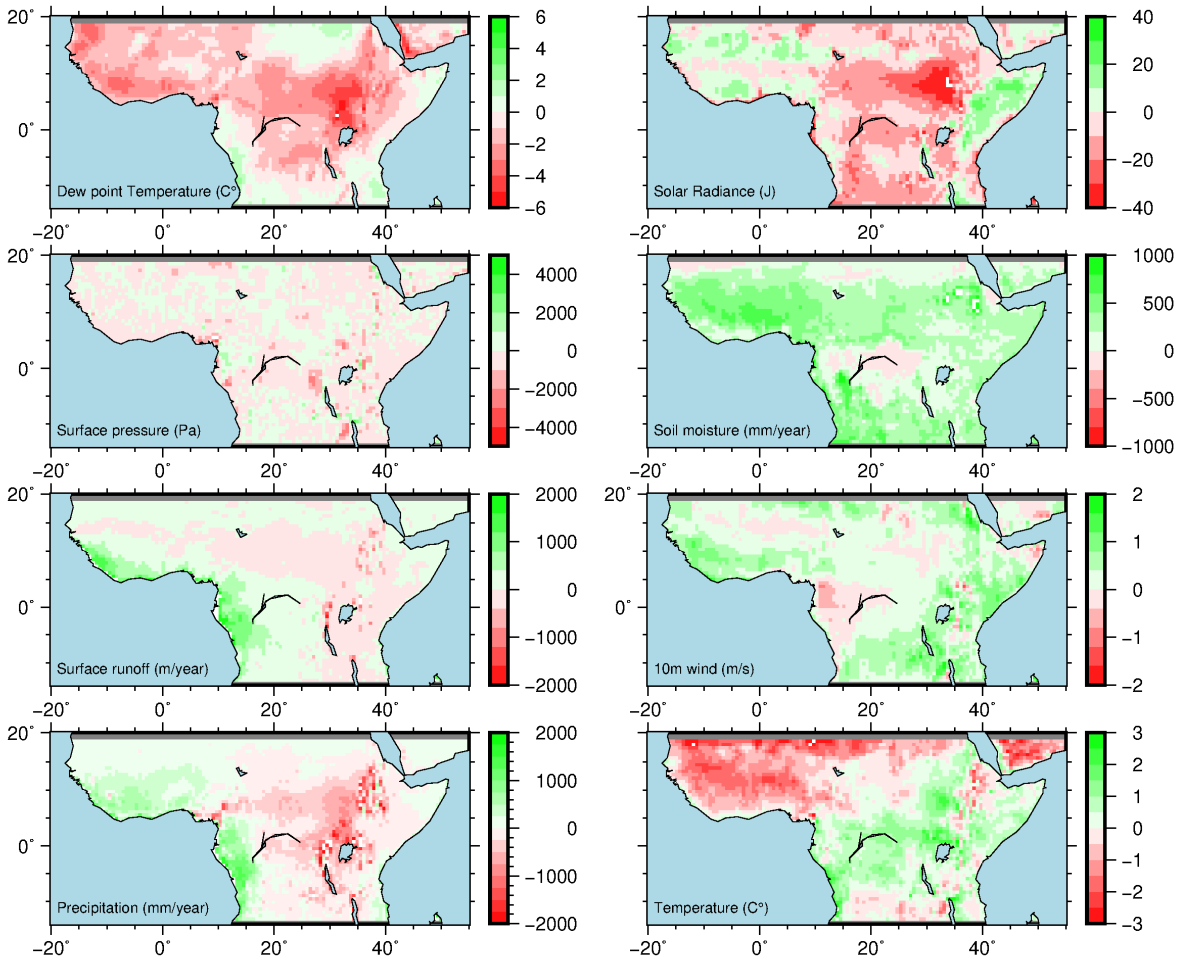

Figure 5.4: Spatial distribution of the difference between the historical REMO and observed ERA5 (REMO – ERA5) for mean temperature, mean precipitation, 10 m wind, dew point temperature, solar radiation, surface pressure, soil moisture and surface runoff.

5.2 Statistical methods

This study is essentially based on applying many statistical methods, which are explained in many statistical textbooks (BAHRENBERG et al., 1999; Coles et al., 2001; Wilks, 2006; Bortz and Schuster, 2011). Many simple and complex statistical concepts are addressed in this chapter: numerical summary measures, time series correlation, statistical inference, linear regression, trend removal and EOF analysis. In each section, the conventional procedure of the statistical method being used is first presented, followed by

its modification and application in the context of this study. All statistical procedures were computed with R statistical software, except for the regression, computed with the scientific programming language Fortran.

5.2.1 Numerical summary measures

Numerical summary measurement is the first step of data preparation and exploration. It is comprised of measures of the location, spread and symmetry. The measures of location used in this study include the mean, the median and the moving average. We used the standard deviation as the measure of spread. The location refers to the data values'central tendency or general magnitude. The arithmetic mean provides an idea of the point which the data addresses, and it is the most fundamental statistical parameter in climatology. The weighted arithmetic mean is preferred to process geographical data – which may have different surface proportions – instead of the simple arithmetic mean. This is because of the convergence of the meridian towards the pole, which makes the grid box of the grid dataset wider at the equator compared to the pole. In this respect, the cosine of the latitude is therefore used as a weighting. The weighting mean is mathematically expressed as the following:

$$x_w = \sum_{i=1}^{n} w_i x_i \tag{5.11}$$

Where w_i refers to the weighting factor, and x_i is the observation. In many applications of this study, it was not necessary to weigh the mean, the w_i is set to 1. The arithmetic mean was mainly used to calculate, to map the climatological mean of the local variables derived from the observational, to simulation datasets, and to present the average crop yield.

To address the strong affect of outliers, in some cases, we used the median instead of the mean. The median is defined as the middle observation and is the point above which half the observations lie and below which half of the observations lie. It is obtained by ordering the observations from smallest to largest and then looking for the middle observation. We used the median to present and map the average maize yield change per grid point because the statistical models did not perform well in some grid points, resulting in very extreme values. The running mean refers to the moving average that we sometimes used. It is a filtering technique used to remove the variations in time series data and compute the means of temporally successive data. In this study, we used it to improve the visualization of some graphs, such as the principal component variable plot

and the future crop yield projection. The moving average is calculated so that an equal amount of data on both sides of a central value is averaged over the selected window. Therefore, the length of the window – which is in our case equal to five – is an odd number. An algorithm has been implemented to compute the running mean.

We calculated the anomaly, based on the arithmetic means, which is the deviation of the element x_i from the mean $\overline{x}$. It could be positive or negative. The anomaly is mathematically expressed as the following:

$$x'_i = x_i - \overline{x} \tag{5.12}$$

The estimation of the climate anomaly is particularly important when the interest is based on the data dynamics and variation rather than its absolute value. They can also be used to detect errors and bias from the datasets. In this study, both the temperature and precipitation empirical orthogonal functions (EOFs) were based on the anomaly datasets.

The spread refers to the degree of variation or dispersion around the centre. We used the variance and the standard variation of the variance measures of the dispersion in the set of data points around the mean. It is calculated as the sum squared of the anomalies and normalized with n, and it is expressed in the following equation:

$$s^2 = \sum_{i=1}^{n} w_i(x_i - x_w) \tag{5.13}$$

In case of equal weighting, the variance becomes the following:

$$s^2 = \overline{{x'_i}^2} \tag{5.14}$$

The variance is always positive. The standard deviation is the measure of the spread most often used, and it is the square root of the variance. The observation is less dispersed (closed to the mean), when the value of the variance/standard deviation is low. In contrast, high values indicate a large dispersion of observations around the mean. The variance was necessary in many statistical methods, for example in the computation of the EOFs.

5.2.2 Person correlation

Person correlation measures the linear association between two variables, which is two time series variables. Supposing that X and Y are two variables, the person correlation is calculated as the following:

$$r_{xy} = \frac{1}{2} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{S_{n,X}} \right) . \left(\frac{Y_i - \overline{Y}}{S_{n,Y}} \right) \tag{5.15}$$

$\overline{X}$ and $\overline{Y}$ refer to the sample means of X and Y, respectively. $S_{n,X}$ and $S_{n,Y}$ denote the sample variance of X and Y, respectively. The correlation coefficient varies between 1 and -1, where 1 denotes a high positive correlation and -1 denotes a high negative correlation. The correlation coefficient of 0 refers to the absence of linear associations between the two variables. In this study, person correlation was used for several reasons, but mostly for measuring the association between principal component predictors used in the crop yield model and large-scale climate interconnection or indices. The objective here was to understand the physical interpretation of each climate principal component, and it was further needed to understand the association between each predictor and maize yield. Person time-series correlation is also embedded in this cross-validated model implementation.

5.2.3 Linear regression and trend removal

Crop yield data are generally characterized by a positive trend due to technological improvements such as irrigation, fertilization, pesticide and improved crops. Trends are traditionally removed to minimize these effects. Like climate predictors, the trend is removed from each of them to deal only with the interannual variation rather than the directional change of climate. Many trend removal approaches exist in the literature, including a second-order polynomial regression model, a moving average model, a locally weighted regression model, a smoothing spline model, an empirical mode decomposition model, the first differential equation (Lobell and Field, 2007) and a simple linear regression method. Among these, the simple linear regression is widely applied. Therefore, we adopted it in this study as the trend removal approach for the crop yield and the climate predictors.

Linear regression is considered the most important and commonly used basic tool in

the quantitative sciences. The aim of linear regression is to create a parametric model for quantitative correlations between different random variables. The univariate variant explains the behaviour of the dependent variable Y, which is here represented by either the time series crop yield or the climate predictors and the independent random variables Xi. The Xi variables are referred to as predictors or regressors. In the special case of single regression, only one predictor is considered by the model. Otherwise, it is multiple regression. To implement linear regression for trend removal, time is considered as the predictor.

$$y_i = a_0 + a_1.x_i + \epsilon_i \tag{5.16}$$

The unknown regression coefficients a_0 and a_1 describe the intercept and slope parameter of the straight linear regression line between X the predictor (time series in our case) and Y the dependant variable (crop yield or climate predictors), respectively. The residual ϵ_i represent the random error and explains the influence of external predictors, which are not included in linear regression. The least-square method is used to estimate these coefficients by minimizing the sum of square of the difference between the observed and predicted. They are estimated as follows:

$$SS\varepsilon = \sum_{i=1}^{n} \left(y_i - \hat{a}_0 - \hat{a}_1 \ .x_i\right)^2 \tag{5.17}$$

The equation is resolved by applying the partial derivative considering the unknown parameters and setting these to zero, as follows:

$$\hat{a}_1 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2} \tag{5.18}$$

The slope $\hat{a}_1$ is estimated by the covariance of X and Y formed by the variance of X, and the intercept $\hat{a}_0$ is obtained by the following equation:

$$\hat{a}_0 = y - \hat{a}_1.\overline{x} \tag{5.19}$$

The detrended data is obtained after subtracting from the crop yield data/climate predictors (y_i) and the predicted values ($ax_i \ + a_0$) according to the following equation:

$$y_i' \ = y_i \ - (a_1 x_i + a_0) \tag{5.20}$$

5.2.4 Non-linear trend and abrupt change analysis

To better interpret the future climate change time series, the non-linear trend and the abrupt change analysis were necessary. The Mann–Kendall test was then applied to the time series for non-linear trend. The Mann–Kendal test is a nonparametric trend test used in climate science. The test was applied to the whole times series (1982 to 2100), including the historical period (1982–2005), the middle-of-the-century period (2020–2060) and the end of the century period (2060–2100). The Mann–Kendall test is formulated when the data X_i of a given time series complies with the following model:

$$X_i = f\left(t_i\right) + \ \varepsilon_i \tag{5.21}$$

Where (t_i) is a continuous monotonic increasing or decreasing function of t and ε_i is the residual time series. Both parameters have the same distribution with a zero mean. The Mann–Kendall test statistic S is given as follows:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn(X_j - X_k) \tag{5.22}$$

Where n represents the length of the time series $X_j \quad ,\ldots, X_n$, $sgn(.)$ is a sign function, X_j and X_k are the values in the years j and k, respectively. The sign function is computed as follows:

$$sgn\left(X_j - X_k\right) = \begin{cases} 1 & if \quad (X_j - X_k) > 0 \\ 0 & if \quad (X_j - X_k) = 0 \\ -1 & if \quad (X_j - X_k) < 0 \end{cases} \tag{5.23}$$

The expected value of S equals zero $[E(S) = 0]$ for the series without a trend, and the variance, $\sigma^2(S)$, are computed as follows:

$$\sigma^2\left(S\right) = \ \frac{1}{18} \ \left[n\left(n-1\right)\left(2n+5\right) - \sum_{p=1}^{q}\left(t_p - 1\right)\left(2t_p + 5\right)\right] \tag{5.24}$$

Where q represents the number of tied groups, and t_p represents the number of data values in the pth group. The test statistic Z is therefore the following:

$$Z = \begin{cases} \frac{S-1}{\sqrt{\sigma^2(S)}} & if \quad S > 0 \\ 0 & if \quad S = 0 \\ \frac{S-1}{\sqrt{\sigma^2(S)}} & if \quad S < 0 \end{cases} \tag{5.25}$$

If $-Z_{1-/2} \leq Z \leq +Z_{1-/2}$, then the null hypothesis of no trend is accepted at the significance level of alpha equal to 0.05. Otherwise, the null hypothesis is rejected, and the alternative hypothesis is accepted at the significance level of alpha.

To determine the approximate year of the beginning of a significant trend, the Sequential Mann–Kendall test is applied. The test is set by Sneyers et al. (1991). The test is based on two series, a progressive one $u(t)$ and a backward one $u'(t)$. The statistically significant trend is confirmed as the two series cross each other and diverge beyond the specific threshold value. The crossing point represents than the year at which the trend begins. The $u(t)$ is normalized at a zero mean and the unit standard deviation, which makes its sequential behaviour to fluctuates around the zero level. In practice, the Mann-Kendall test considers the relative values of all terms in the times series and is computed according to the following sequence:

- The magnitudes of x_j annual mean time series ($j = 1, \ldots, n$) are compared with x_k. At every comparison, the number of cases is counted and denoted by n_j

- The test statistic t is given by the following equation:

$$t = \sum_{1}^{j} n_j \tag{5.26}$$

- The mean and variance of the statistic are estimated as follows:

$$e\,(t) = \frac{n(n-1)}{4} \tag{5.27}$$

$$var\ t_j = \frac{j(j-1)(2j+5)}{72} \tag{5.28}$$

The sequential values of the statistic u are approximated as follows:

$$u\,(t) = \frac{t_j - e_{(t)}}{\sqrt{var(t_j)}} \tag{5.29}$$

A similar sequence is applied to compute $u^{'}\,(t)$ but starting from the end of the series.

5.2.5 EOF analysis

The EOF, also known as Principal Component Analysis (PCA), is a multivariate technique used to reduce the dimensions of large datasets by creating a set of variables called principal components (PCs). The PCs are built in such a way that they are orthogonal from one another and according to the assumption that the total variance of the original data can be completely reproduced by a linear combination of PCs and corresponding to a fixed weight called eigenvectors. The EOF is particularly useful for highly correlated variables containing redundant information. Many scientists have documented its importance in meteorology and climatology (Jolliffe, 2002). Besides using a multivariate technique, the EOF is also applied to different spatial units of a single variable to generate large-scale spatial patterns of centres with modes of variation, as explained by von Storch (1999). In this study, the EOF analysis was first applied to spatial units (time series gridded points) of a single variable. It was applied to two variables separately: mean precipitation and mean temperature. Another EOF analysis was applied to PCs obtained from the precipitation and temperature spatial units.

It is important to define the dataset and explain some basic operations before explaining how the EOF analysis works. A data matrix X of a certain climatic field variable, for example, the mean times series precipitation of January derived from the CRU dataset, has a dimension of $(t \times m)$ with $j = 1, m$ spatial units over $i = 1, t$ time steps. With a spatial resolution of 0.5°, the number of grid points – which represent the number of variables in EOF analysis – is equal to 4346 ($m = 4,346$), and the time step t is equal to 57 (from 1960 to 2016). The EOF analysis is generally applied to anomaly datasets denoted by $X^{'}$. The anomaly $X^{'}$ is built by subtracting from X the corresponding temporal means, as follows:

$$X'\ (t,s) = X\,(t,s) - \overline{X}(s) \tag{5.30}$$

With $\overline{X}\,(s)$ as the arithmetic means per grid point. Another operation before the computation of the EOF is the weighting. The datasets were weighted because of the non-

uniformity over the surface of the earth (denser poleward). The non-uniformity implies that the number of grid points per unit area increases with increasing latitude because the meridians converge towards the poles. Therefore, the PCA for this kind of gridded data will dominate in high-latitude features. We weighted each data point by the local area of its location – this approach is widely used in scientific communities. We multiplied each gridded point by the cosine of its latitude as follows:

$$w(s) = \sqrt{\frac{cos(s)}{\sum_{s=1}^{n} cos(s)}} \tag{5.31}$$

After this step, the principal component times series can be reconstructed as follows:

$$X(t,s) = \sum_{i=1}^{m} c_i(t) v_i(s) \tag{5.32}$$

Where $c_i(t)$ refers to the principal components (PCs), which are the linear combination of the original datasets $X(t,s)$. The variance of $X(t,s)$ decreases with increasing index $v_i(s)$ is its associated pattern or eigenvectors, m refers to the maximum number of variables and t is the time. The EOF can be performed on either a correlation matrix or a covariance matrix. The gridded points of the EOF were performed on the covariance matrix to allow the largest possible variations in the dataset. The covariance matrix is therefore written as follows:

$$\sum = n^{-1} X'^{T} X' \tag{5.33}$$

Before the obtention of the derivation of PCs, a main mathematic problem should be solved. This is how to find the eigen vector $v_i(s)$ and the associated eigenvalues λ_i . This is presented in the equation formulated by Von Storch and Zwiers (2004) as follows:

$$A \times E = \lambda \times E \tag{5.34}$$

The dimensions of A and E are $(m \times m)$ and $(m \times 1)$, respectively. With a dataset of m grid points, the principal component equation should be resolved m times. In the first step, the covariance matrix of the data ($A\ X'^{T} X'$) must be solved, and the first eigenvector $v_1(s)$ and its associated eigenvalue λ_1 are obtained. The first eigenvector contains the largest variance. This process will be repeated m times, and A is computed each time after subtracting the information explained by the i^{th} eigenvector and the next eigenvector. All eigenvectors are uncorrelated. The proportion of the variance explained

by each eigenvector is therefore described in the following equation:

$$Var(i) = 100.\frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} \tag{5.35}$$

The variance explained by each eigenvector decreases as the index of the eigenvectors increases, as illustrated in Figure 5.5. The eigenvectors obtained are further normalized and sorted. The normalization implies the following:

$$\sum_{s=1}^{m} v_i^2 (s) = 1 \tag{5.36}$$

Figure 5.5 displays the explained variance in the percentage of the 50 leading empirical orthogonal functions of each month for the precipitation and the temperature. The first EOF of the temperature for every month contains more information of the total variables (almost 45 % of the total variance explained, on average), while the second EOF explains only 18 %. This is not the case for the precipitation EOFs, where the first leading EOF contains more information of the whole dataset (15 % of the explained variance, on average) but close to the second (11 % of the explained variance). The explained variance of the precipitation and temperature decreases progressively with the number of EOFs. This difference led us to take only the two leading EOFs of the temperature for each month, which explains, on average, 60 % of the total variance. For the precipitation, we took the five leading EOFs.

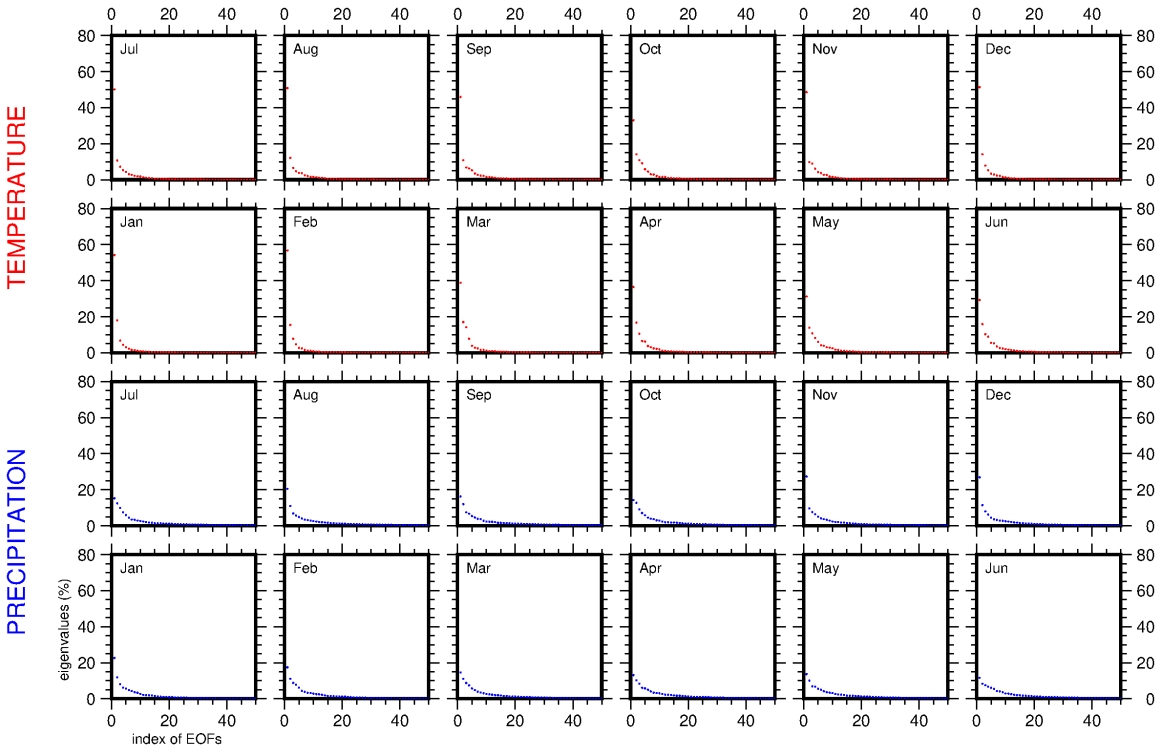

Figure 5.5: Variance explained by each of the leading 50 EOFs of precipitation and temperture

So far, the EOF analysis is performed on the observational dataset, CRU. To derive simulated principal components, the anomaly simulated REMO datasets were projected into the spatial pattern of the CRU dataset $v_i(s)$ according to the following equations:

$$u_i(t) = \sum_{s=1}^{m} v_i(s) \times X''(t,s) \tag{5.37}$$

With $u_i(t)$ as the simulated REMO PCs. These operations are also applied to the remaining eleven months of the precipitation dataset and to the twelve months of the temperature dataset of both observational the CRU dataset and simulated REMO. The five leading PCs of precipitation datasets for every month and the two leading PCs of temperature datasets for every month were retained for further EOF analysis. These leading PCs retained the mode of climate variability in the study area and the remaining noises. We therefore computed a second EOF analysis on these 84 PCs, 60 from the precipitation datasets and 24 from the temperature datasets. In contrast to the gridded

EOF analysis as described bellowed, the EOF analysis applied to the 84 PCs is built on the correlation matrix because it comes from different datasets (precipitation and temperature), which have different units and variability. The same operation is applied to the simulated REMO PCs. The final leading PCs obtained serve as a predictor of maize yield.

6. Specific crop (Maize) water balance model estimation

This chapter describes the estimation of the processed-based crop model, known as the WRSI. The WRSI is an indicator used to monitor weather-related food insecurity, mainly in developing countries where rain-fed agriculture prevails (Turner, 2020). The computation of WRSI is simple and requires the precipitation, the crop coefficient and the reference evapotranspiration. But in this study, we used the most advanced version, which uses the actual evapotranspiration instead of the precipitation. The WRSI estimates the water supply and demand of any crop during the growing season. Therefore, it is considered as a simple process-based crop model and widely used in Famine Early Warning Systems to inform food security and as a management instrument of crop insurance. The WRSI is directly related to agriculture drought in contrast to other drought indices, such as the standardized precipitation index (Dlamini, 2013), the dry spells assessments (Masupha and Moeletsi, 2018) and the Standardized Precipitation Evapotranspiration Index (Vicente-Serrano et al., 2015), which all refer to meteorological drought and do not take into consideration the crop specificities and the soil water. The Normalized Difference Vegetation Index is also used to monitor drought, but it is not designed for specific crops like WRSI. Agriculture drought is characterized by a soil water deficit and a low-value WRSI (Mishra and Singh, 2010), and can be perceived in any development phase of the crop, including root development, leaf and stem development, flowering, pollination, grain filling and grain quality (Prasad et al., 2008). Legesse (2010) classified the values of the WRSI for drought conditions, and crop performance is indicated in Table 6.1.

Many studies have demonstrated that the WRSI is strongly correlated to crop yields (Doorenbos and Pruitt, 1977; Frere and Popov, 1979; Frère and Popov, 1986). It can therefore be used as a valuable proxy of crop yield, especially in areas where crop statistics are missing. Verdin and Klaver (2002) reported a correlation coefficient of 0.75 between WRSI and crop yield. Similarly,Tarnavsky et al. (2018) reported that WRSI could explain

Table 6.1: Classification of WRSI according to drought conditions and crop performance (FAO, 1986; Legesse, 2010)

WRSI	Drought classification	Crop performance description
80 – 100	No drought	Good
70 – 79	Mild	Satisfactory
60 – 69	Moderate	Average
50 – 59	Severe	Poor
50 <	Extreme	Total crop failure

up to 62 % of yield variability in Tanzania. Senay and Verdin (2003) also found a good correlation between the WRSI and maize production in Ethiopia. Verdin and Klaver (2002) reported a good performance of the WRSI in predicting crop yield in Zimbabwe. Many other studies have been conducted to forecast maize yields by the WRSI around the world (Martin et al., 2000; Rafi and Ahmad, 2005; Jayanthi et al., 2013). It is important to note that the strength of that correlation varies according to crop and location (Jayanthi et al., 2013). Some studies used future climate simulations to assess the impact of climate change on maize-related WRSI (Crespo et al., 2011; Masupha and Moeletsi, 2018).

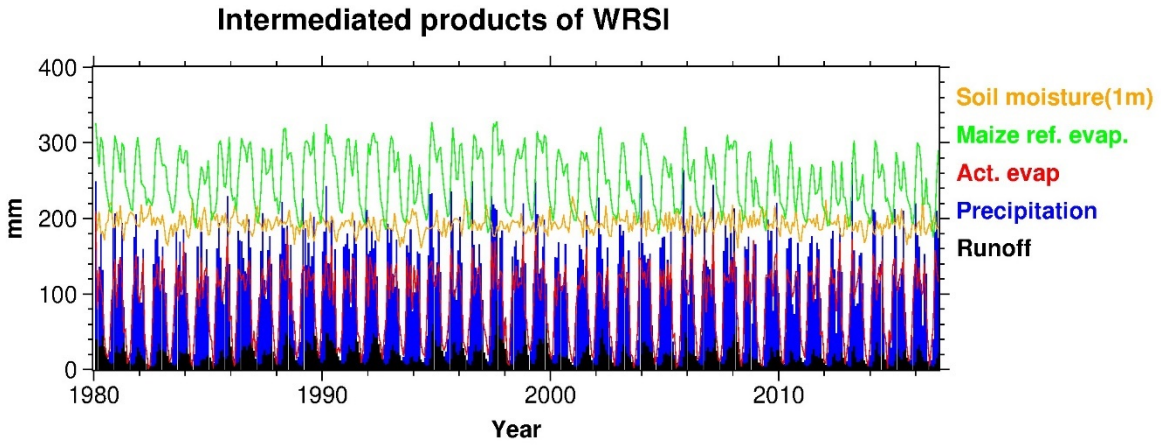

Figure 6.1: Monthly variability of the soil moisture, reference evapotranspiration, actual evapotranspiration, precipitation and runoff.

The WRSI is obtained after estimating some intermediated parameters: the reference evapotranspiration, the crop water requirement and the actual evapotranspiration. Figure 6.1 presents the temporal variability of the soil moisture, maize reference evapotranspi-

ration, actual evapotranspiration, precipitation and surface runoff of the city of Kinshasa (Capital of the Democratic Republic of the Congo). Some variables (surface runoff, precipitation and soil moisture) have been directly downloaded. Others were calculated:the actual evapotranspiration and the maize reference evapotranspiration, was calculated .

In this study, the WRSI is first used as a stand-alone proxy of crop yield performance and then as a predictor of the maize yield statistical model.

6.1 Reference evapotranspiration

The concept of reference evapotranspiration was first used by irrigation researchers and engineers in the late 1970s (Allen et al., 1998). It is was introduced in contrast with potential evapotranspiration, but both are very closed and sometimes used interchangeably. Reference evapotranspiration is defined as "the rate of evapotranspiration from a hypothetical reference crop with an assumed crop height of 0.12 m, a fixed surface resistance of 70 sm^{-1} and an albedo of 0.23, closely resembling the evapotranspiration from an extensive surface of green grass of uniform height, actively growing, well-watered, and completely shading the ground". The hypothetical crop is free of disease and water stress. In contrast, with the potential evapotranspiration, the rate is not based on a specific crop. Instead, many savannah shrubs and agricultural crops are concerned (Allen et al., 1998).

There are several ways to estimate crop reference evapotranspiration in the literature. However, the lysimeter approach is the most accurate one, which also served to develop and validate other methods (Allen et al., 2011). Because of the complexity and financial cost, the lysimeter approach is constrained to only fundamental research (Allen et al., 2011). Thus, mathematical models that use meteorological data are commonly used for application purposes. In this study, we used the most extensive combined method approved by the Food and Agricultural Organisation (Allen et al., 1998; McVicar et al., 2007), which is also known as the Penman–Monteith equation. The model uses the standard climatological records of solar radiation (sunshine), air temperature, humidity and wind speed for daily, weekly, ten-day or monthly calculations. We calculated the crop reference evapotranspiration for each growing season based on monthly data (the initial stage, development, the mid-season stage and late season). The length of the different stages of the growing season, according to the region is defined in Table 6.2:

- The initial period corresponds to the germination, and during this phase, the soil

is naturally dry and uncovered.

- The ground cover is around 70 to 80 % during the crop development stage, and evapotranspiration becomes the main process.
- The maximum of groundcover is reached during the mid-season stage, which is also known as the maturity stage of the crop. This stage is also the longest for most annual and perennial crops (40 *days* for maize growing in humid regions like Nigeria).
- The late season is the last stage, starting from the crop's maturity and lasting to the senescence or harvest.

This used the calculation applied in Nigeria and east Africa. The reference evapotranspiration is mathematically expressed as such:

$$ET = \frac{0.408\Delta\,(Rn - G_{sfc}) + \gamma\frac{900}{T+273}\,u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{6.1}$$

$$e^0\,(T) = 0.6018exp(\frac{17.27T}{T + 237.3}) \tag{6.2}$$

$$e_s = \frac{e^0\,(T_{max}) + e^0(T_{min})}{2} \tag{6.3}$$

$$e_a = \frac{e^0\,(T_{max})\,\frac{RH_{min}}{100} + e^0(T_{min})\frac{RH_{max}}{100}}{2} \tag{6.4}$$

$$\Delta = \frac{4098\left[0.6108\ exp(\frac{17.27T}{T+237.3})\right]}{(T + 237.3)^2} \tag{6.5}$$

$$\gamma = 0.665x10^{-3}\ P \tag{6.6}$$

Where ET refers to the reference evapotranspiration ($mm \cdot d^{-1}$), Rn corresponds to the net radiation ($MJ \cdot m^2 \cdot d^{-1}$), G is the soil heat flux in the ground ($MJ \cdot m^2 \cdot d^{-1}$), e_s is the saturation vapour pressure of the air temperature (KPa) and e_a is the actual vapour pressure (KPa). T is the main daily temperature ($^\circ C$), u^2 wind velocity is at 2 m height ($m.s^{-1}$), Δ is the slope of the saturation vapour-pressure curve of the air temperature ($KPa \cdot^\circ C^{-1}$), RH is the relative humidity, P is the air pressure (KPa); $e_s - e_a$ is the saturation vapor pressure deficit (kPa) and γ is the psychrometric constant ($kPa^\circ C^{-1}$)

Figure 6.2 illustrates the climatological mean (1982–2016) of the seasonal spatial distribution of the reference evapotranspiration in tropical Africa from the ERA5 dataset.

Independent of the month, the reference evapotranspiration over tropical Africa increases northward and southward from the Guinean coast and the Congo Basin with the increasing latitude. The mean reference evapotranspiration reaches the minimum values of about 200 $mm/month$ in the Congo Basin and along the Guinea Coast. Maximum values of around 500 $mm/month$ are recorded in the northern region around Sahel. A meridional belt of high values from the west of Sahel to the Somalian coast can be observed, particularly during January, February, March and December. Another strong longitudinal pattern of high values is observed in the Congo Basin from July to October. This spatial pattern observed in Figure 6.2 corresponds to the finding of Abiye et al. (2019), who used the potential evapotranspiration product of the CRU dataset. According to Abiye et al. (2019), the spatial distribution of the potential evapotranspiration over Africa is highly influenced by the cloud cover, which reduces the solar radiance. The characteristics of the local conditions, such as topography, vegetative cover, availability of soil moisture, reflective land surface, and changes in land use/land cover also influence the reference evapotranspiration (Rim, 2009; Ashaolu and Iroye, 2018).

Table 6.2: Lengths of Maize development stages and climatic regions (days) (Allen et al.,1998).

Init.	**Dev.**	**Mid.**	**Lat.**	**Total**	**Plant Date**	**Region**
30	50	60	40	180	April	East Africa (alt.)
25	40	45	30	140	Dec./Jan.	Arid Climate
20	35	40	30	125	June	Nigeria (humid)
20	35	40	30	125	October	India (dry, cool)
30	40	50	30	150	April	Spain (spr, sum.); Calif.
30	40	50	50	170	April	Idaho, USA

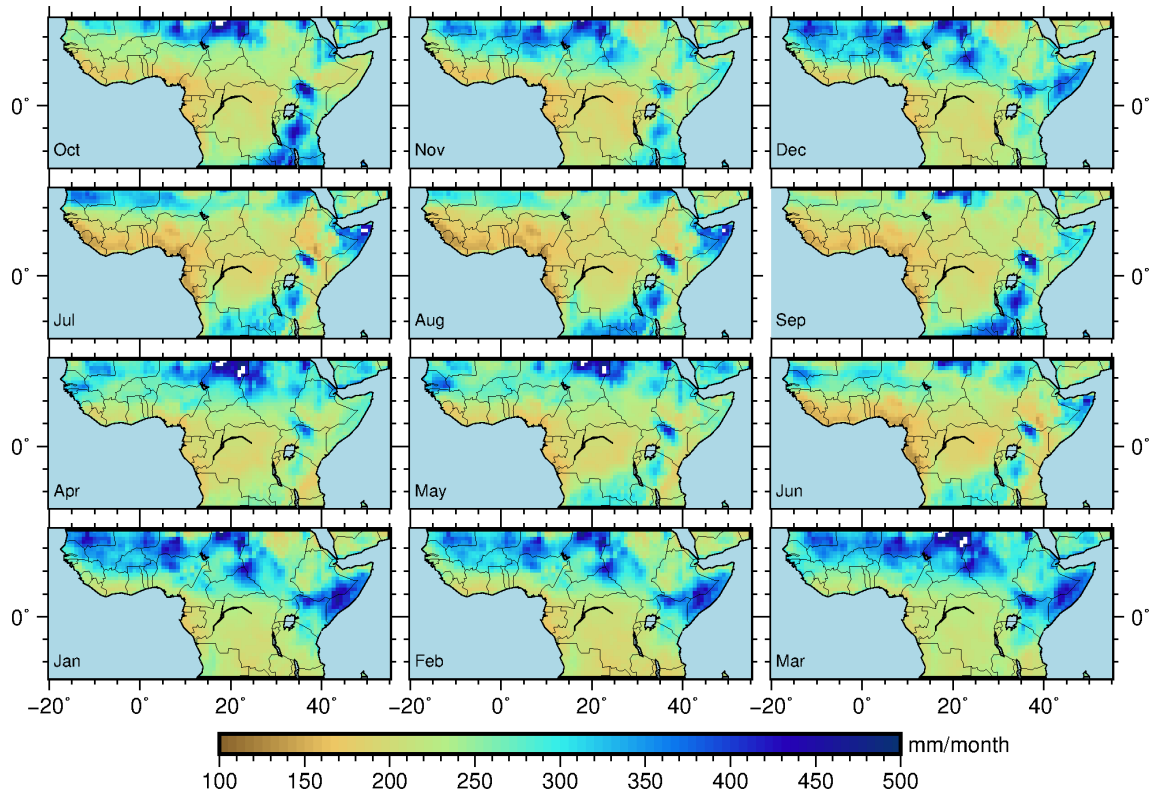

Figure 6.2: Spatial pattern of the monthly reference evapotranspiration (1982–2016).

6.2 Crop water requirement

The crop water requirement is a product of the reference evapotranspiration (ET) and the specific crop coefficient (Kc), maize in our case. It is calculated for every growing season, as the following:

$$WR = K_c \times ET \tag{6.7}$$

The crop coefficient indicates the pattern of water used by each crop during different growing seasons (the initial stage, development, the mid-season stage and late season). The methodology and framework used to estimate it has been extensively explained by the Food and Agriculture Organization Irrigation and Drainage Paper 56 (Allen et al., 1998). In addition, the FAO provides a look-up table of the values of the crop coefficient for a wide range of crops at different growing seasons. The crop coefficient curve for maize is displayed in Figure 6.3. In the initial phase, the crop coefficient is very low because the ground cover is less than 10 %. The maximum of the Kc reaches the maximum value

and remains constant in the mid-season stage. It drops again in the late season.

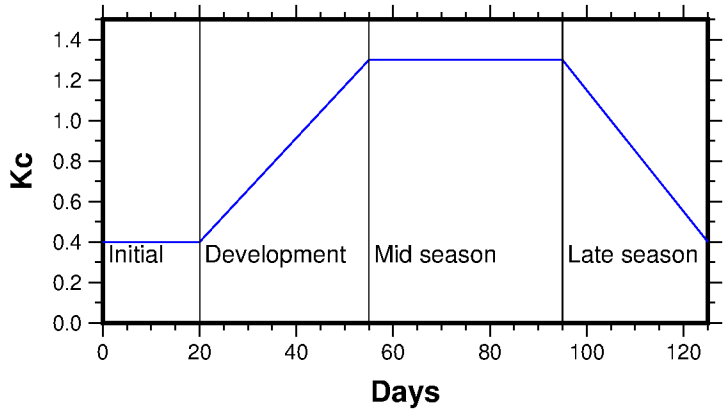

Figure 6.3: Crop coefficient curve for Maize (Allen, 1998)

6.3 Actual evapotranspiration

The actual evapotranspiration represents the actual amount of water lost from the soil water reservoir that has become atmospheric water vapour under natural conditions (Xu et al., 2006). Many approaches exist in the literature to estimate the actual evapotranspiration from meteorological observation and satellite devices. In our study, we used the rained-driven water balanced accounting shame approach, which is based on the water cycle (Campos et al., 2016); it is easy to implement. The soil water balance equation is expressed as follows:

$$AET = R + I \ \pm \Delta SM - D_p - RO + CR \tag{6.8}$$

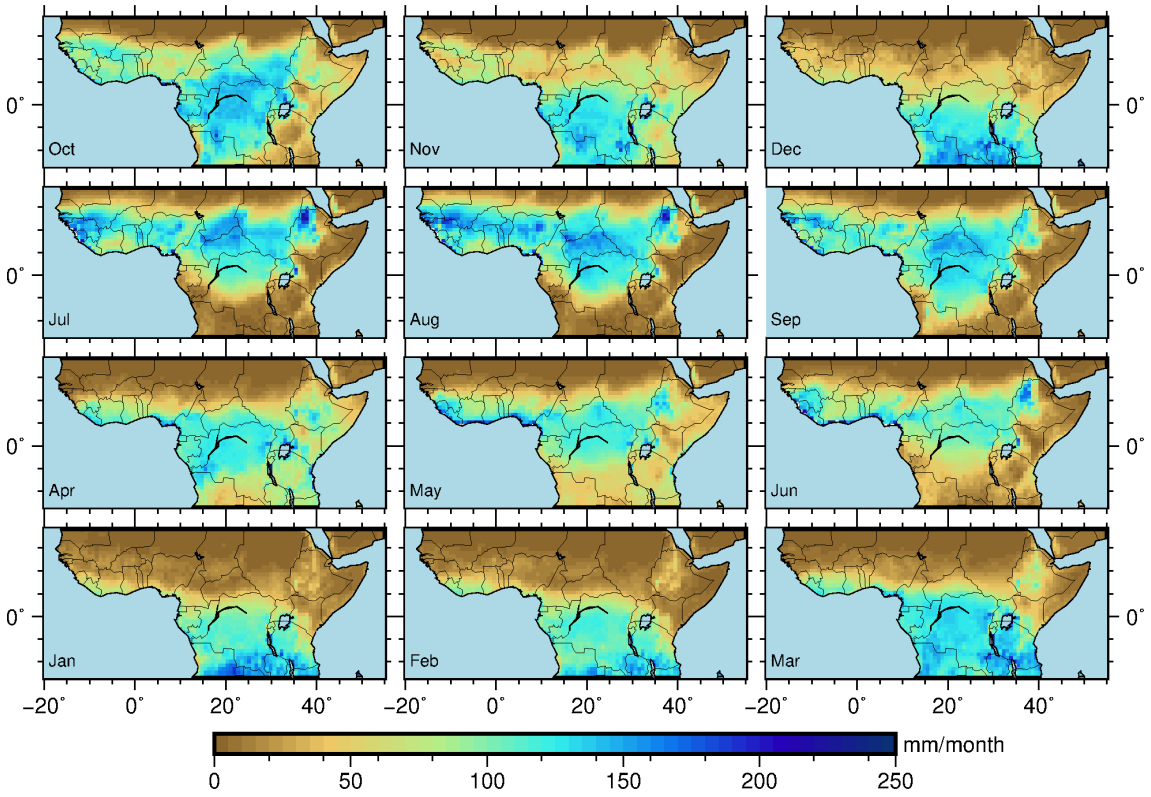

Figure 6.4: Spatial pattern of the monthly actual evapotranspiration (1982-2016).

Where AET is the actual evapotranspiration; ΔSM is the change of water content between two consecutive days, months in our case; I is the irrigation, measured in mm; D_p

is the deep percolation in mm; R is the precipitation in mm; RO is surface runoff flux in mm and CR is the capillary rise flux in mm. The irrigation term was neglected because we assume a red-fed agriculture. The drainage term was also neglected because we are only concerned with the top layer soil water content (90 cm) as well as the capillary rise flux due to no contribution from groundwater with a capillary rise into the root zone (Ridolfi et al., 2008).

Figure 6.4 presents the spatial pattern of the actual monthly evapotranspiration of the study area. Generally, the actual evapotranspiration decreases northward and southward from the Congo Basin with increasing latitude. Low values are observed in arid and semi-arid regions, excepting months of October, November, December, January, February and March, when high values are observed south of the Congo Basin.

6.4 WRSI

The WRSI is simply a ratio between the actual evapotranspiration and the potential evapotranspiration, as expressed in the following equation:

$$WRSI = \frac{WR \times 100}{AET} \tag{6.9}$$

Figure 6.5 illustrates the time series of the WRSI of some cities in Africa. A strong interannual variability of the WRSI is observed in Porto Novo in western Africa. In the heart of the Congo Basin, the index is high but has a weak inter-annual variability (Kisangani). The WRSI means in Johannesburg (Southern Africa) and Kigali (East Africa) are relatively low throughout the years, requiring additional water supplies for good crop production.

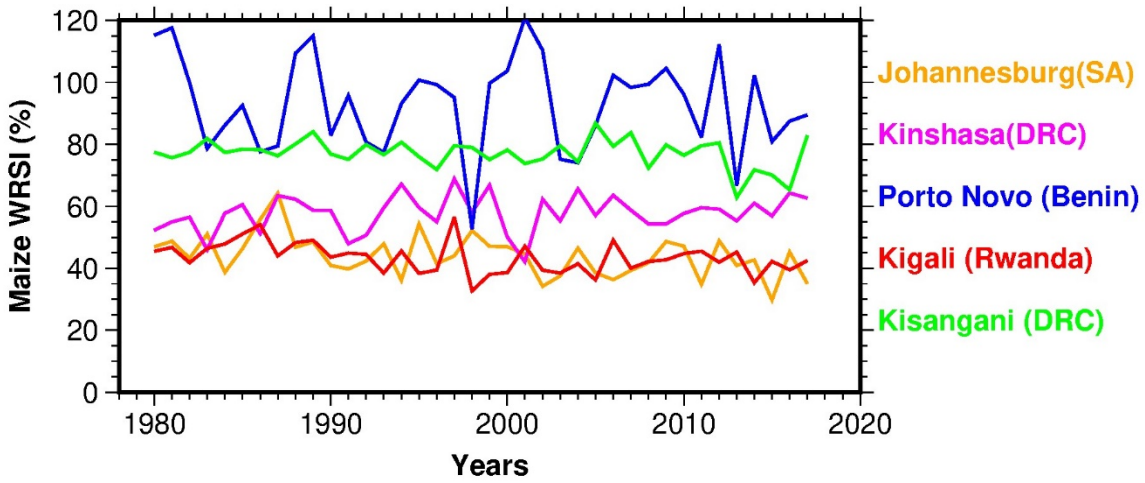

Figure 6.5: Inter-annual variability of WRSI in Johannesburg, Kinshasa, Porto Novo, Kigali and Kisangani.

7. Development of the statistical crop yield model

Various statistic regressions have been used to model the impact of climate change on crop yields. They are all assessed according to their ability to deal with multi-collinearity, non-climatic trend removal, selection of relevant predictors and their ability to better fit the validation dataset. According to Lobell and Burke (2009), collinearity is one of the most important challenges to resolve in modelling the impact of climate change crop yields because it leads to model misspecification. For example, it is challenging to assess the effect of the mean temperature, maximum temperature and precipitation in statistical models (Shi and Tao, 2014). Sharif et al. (2017) compared the performance of six regression techniques in predicting oilseed rape yield response to climate variability. Their result indicated that Ordinary Least Squared (OLS) multiple regression and stepwise linear regression were less effective, whereas Lasso and Elastic Net regressions were the most effective. Ridge regression and the partial least square show intermediated results. Despite this result from other researchers, OLS multiple linear regression is still commonly used because of its simplicity and ability to integrate non-linear variables (Chen et al., 2011; Lobell and Burke, 2010; Lobell and Field, 2007; Schlenker and Lobell, 2010). However, it requires deep biological knowledge of the variables being used to avoid collinearity issues and more attention needs to be paid to the number of predictors to avoid the overfitting problem. In this study, we suggested a cross-validated version of OLS multiple regression, developed by Paeth and Hense (2003) and recently improved by (Awoye et al., 2017). This regression handles large datasets (highly correlated) well and is less prone to overfitting. In addition, the model is performant in capturing physically based predictors. It is applied in two phases: (1) The first phase involves of developing or calibrating the model. It combined many statistical techniques, such as forward stepwise multiple linear regression, cross-validation, bootstrapping and model weighting; (2) The second phase, the forecast phase, consisted of applying the equation derived from the development phase to the climate-simulated predictors.

7.1 Cross-validated multiple regression

The multiple crossed validated regression aims to assess the impact of predictors on the predictand and to model the future maize yield changes. The regression is applied to each bootstrap sample for every grid point. The selected predictors time series are assembled in the OLS linear regression model, as follows:

$$y_i \;=\; a_0 \;+\; \sum_{j=1}^{k} a_j x_{ij} + \varepsilon_i \tag{7.1}$$

Where the crop yield (predictand) y_i at year i is explained by a linear combination of k predictors x_{ij}, and ε is the random error or the unpredictable component of the predictand. In the matrix format, the same equation can be written as follows:

$$y = \; \boldsymbol{X}\overline{a} + \epsilon \tag{7.2}$$

Where y is the predictand of one dimension, containing n observations, and X is a two-dimension matrix, representing the k predictors and n observations. The $(k+1)$-dimensional vector a represents the intercept, and the coefficients corresponding to the k predictors and ϵ represent the random errors of each observation. The intercept and coefficients a are estimated using the least square approach (OLS), which minimizes the sum of square errors, according to the following mathematical expression:

$$a \;=\; (XX)^{-1} X^T\, y \tag{7.3}$$

a is comprised of the intercept and parameters of the predictors a_j.y contains the predictand times series with $i = 1 \ldots n$. X represents a matrix with the value 1 in the first column and the predictors time series in the following column. The explained variance of the model of the multiple linear regression is calculated as the following:

$$R^2 \;=\; y^T (X(X^T X)^{-1} X^T \;-\; U)\, y \left[y^T (I - U)\, y\right]^{-1} \tag{7.4}$$

Where I represent the $n \times n$ identity matrix and U an $n \times n$ matrix with all entries equal to $1/n$. The distribution of the explained variance R_2 under the null hypothesis H_0 is Fisher F, and its significance level can be estimated given a certain error level.

The multiple linear regression explains how different predictors could influence the pre-

dictand but neither ranks the contribution of each predictor to the model nor selects the most important predictors. This is possible only through a stepwise screening approach. The stepwise linear regression can use two different algorithms: the forward or backward selection. In this study, we applied the forward selection: predictors were added to the regression model step by step according to their order of importance. The model starts by taking the mean of the mean of the predictand y as intercept, which is calculated as follows:

$$a_0 = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{7.5}$$

At this first step of the model, the X matrix only contains the first column – an n dimensional vector with a value equal to 1. The most important predictor is selected in the second step, after the linear correlation – described in subsection 5.2.2 – between to predictand and each predictor. The predictor holding the highest correlation coefficient with the predictand is the one selected. Before the third step, the simple OLS regression is computed based on the predictor selected in the second step and the predictand, according to equation 5.16. In the third step, the second most important predictor is chosen based on the linear correlation between the remaining predictors $(k-1)$ and the residual predictand times series derived from the simple linear regression computed in the second step; the predictor holding the highest coefficient with the residual times series is selected as the second most important predictor of the model. Instead of the simple linear regression, the multiple regression is computed with the two predictors selected in previous steps. The third most important predictors will be chosen in the fourth step, after the linear correlation between the residual time series of the multiple linear regression and the remaining predictors $(k-2)$. From the fourth step, the model will iterate until the last predictor is selected.

7.2 Cross validation, bootstrap sampling and variable selection

To avoid spurious correlation between the residual time series and predictors, as explained in stepwise linear regression, a cross-validation analysis was performed by splitting the datasets into two parts: calibration and validation. The splitting process was done by the bootstrap sampling procedure. The Bootstrap sampling procedure draws sample permutation. It combines the uniform random-number generators and the kernel density

estimate. It consisted of choosing an integer number and updating and rescaling the seed integer number to the interval between 0 and 1. We proceeded by applying the congruential generator algorithm defined as follows:

$$S_n \ = \ a \ S_{n-1} \ + c, \ Mod \ M \tag{7.6}$$

$$u_n \ = \ S_n/M \tag{7.7}$$

Where S_{n-1} refers to the seed of the previous iteration, S_n is the updated seed and a, c and M are the integer parameters called the multiplier, increment and modulus, respectively. The quantity u_n is the uniform variate produced by the iteration defined by the estimation of the center of the kernel according to the following equations (algorithm):

$$choose \ x_i \ if \ \frac{i-1}{n} \leq u < \frac{i}{n}, \tag{7.8}$$

Which yields

$$i = int\,[nu + 1] \tag{7.9}$$

With $[*\ ]$ indicating the retention of the integer part only. With the splitting of the dataset, the equation 7.1 is computed only in the calibration datasets. In each step of the stepwise linear regression, the mean square error of the model (MSE_M) is computed as follows:

$$\widehat{MSE_M} = \ \frac{1}{n-nb} \sum_{j=1}^{n-nb} (y_j \ - \ \widehat{y}_j)^2 \tag{7.10}$$

Where nb refers to the number of observations, commonly called "bootstrap elements" in the validation dataset. y_j indicates the value of the calibration datasets, entering the regression model at time j, and $\widehat{y}_j$ is the estimated predictand derived from the regression model. Until now, the predictors are ranked and ready selected, and the MSE_M always decreases when a predictor is added to the model. To ensure the selection of the important variables – which has a physical interpretation – the mean square error of the calibration dataset (MSE_{CV}) is calculated as follows:

$$\widehat{MSE_C} = \ \frac{1}{nb} \sum_{j=1}^{nb} (y_j^B \ - \ \hat{y}_j^B)^2 \tag{7.11}$$

Where y_j^B refers to the bootstrap elements of the calibration dataset, and $\hat{y}_j^B$ is the estimated predictand of the regression equation, derived from the calibration dataset but

applied to the validation dataset.

Compared to MSE_M, the MSE_C decreases when a relevant predictor (physical meaning) is selected. An artifact predictor is likely to increase the MSE_{CV}. Thus, the stepwise multiple regression stops in the first relative minimum value of MSE_{CV}, and the optimal number of predictors are selected and ranked according to their importance. The whole process of model development described above was iterated 100 times to ensure the robustness of the model, which resulted in 100 randomized bootstrap samples.

In this study, the multiple cross-validated regression model is used to model the impact of climate change on maize yield. The model is applied in each grid box, which has a spatial resolution of 0.5°. The list of predictors is comprised of the four leading PCs derived from the second EOF analysis: the mean precipitation, the maximum temperature, the minimum temperature, the mean temperature, and the five WRSI products. As far as the local predictors are concerned, for a given grid box, we calculated the average of the grid point concerned with its eight neighboring grid points to avoid a potential spatial offset error of the local predictors. The dataset covers the period from 1982 to 2016 (34 observations). Only 28 observations were used to calibrate the model and six to validate it ($nb = 6$). The performance of the single or multi-averaging models was assessed by the means of the coefficient of determination and the root mean squared error of the model.

7.3 Model weighting

The bootstrap procedure in each grid point generates 100 samples, which means 100 sub-regression models with an optimal selection of predictors. In a normal procedure, the final estimated predictand values should derive from the average over the 100 sub-models. However, in this study, we used six different multi-model averaging methods: (1) simple averaging; (2) weighting with the adjusted coefficient of determination of the individual models with respect to the observational data; (3) the Akaike weights; (4) the corrected Akaike weights; (5) Bayesian information criteria; and (6) Bayesian model averaging. Awoye et al. (2017) compared the impact of different model averaging in predicting maize yield. Their results indicated that the automated Bayesian model (not used here) was the best, followed by the Bayesian model average and the Bayesian information criteria. The bootstrap procedure serves to further define a robust confidential estimate around the multi-model mean, which is important to assess the uncertainties of the model. At

each time step, the confidential interval is calculated as follows:

$$CI_{95\%} = (\overline{Y}_w - 1.96 \times S_{Wn,y}; \overline{Y}_w + 1.96 \times S_{Wn,y}) \tag{7.12}$$

Where $\overline{Y}_w$ and the $S_{Wn,y}$ denotes the average yield or average yield changes and its corresponding standard deviation, respectively. The standard deviation corresponds to the standard error, which was computed on 100 samples. Therefore, 1.96 corresponds to the 97.5 percentile of the standard normal distribution.

7.3.1 Simple averaging

The simple averaging weighting is similar to the arithmetic mean, computed in equation 5.4 , the formula is here expressed as follows:

$$E(\Delta) \; = \; \frac{1}{2} \; \sum_{i=1}^{B} \Delta_i \tag{7.13}$$

7.3.2 Adjusted coefficient of determination

The averaging of the coefficient of determination of every model iteration can be computed as follows:

$$R^2_{(i)} = \frac{\left[\sum_{j=1}^{n} \left(Y_j \; - \overline{Y}\right) \left(f_j^{(i)} - \; \overline{f}^{(i)}\right)\right]^2}{\left[\sum_{j=1}^{n} \left(Y_j \; - \; \overline{Y}\right)^2\right] x \left[\sum_{j=1}^{n} \left(f_j^{(i)} \; - \; \overline{f}^{(i)}\right)^2\right]} \tag{7.14}$$

Where Y_j refers to the observed yield data and $\overline{Y}$ the mean value of the observed yields. $f_j^{(i)}$ is an estimate of the model M_i at time j, and $\overline{f}^{(i)}$ represents an estimate of the average yield from the mode M_i in the reference period. The major issue with the averaging of the coefficient of determination is that it increases every time a new predictor is added, limiting a good comparison of models which have different numbers of predictors. It therefore needs to be adjusted. The adjusted coefficient of determination is expressed as follows:

$$\overline{R}^2_{(i)} \; = \; 1 - \frac{(1 - R^2_{(i)})(n-1)}{n \; - \; p \; - \; 1} \tag{7.15}$$

Where p refers to the number of free parameters in the model $M_{i.}$ The quantity of interest is estimated as follows:

$$E(\Delta) \;=\; \frac{\sum_{i=1}^{B} \overline{R}^2_{(i)} \times \; \Delta_i}{\sum_{i=1}^{B} \overline{R}^2_{(i)}} \tag{7.16}$$

Where p refers to the number of free parameters in the model M_i.The quantity of interest is estimated as follows:

$$E(\Delta) \;=\; \frac{\sum_{i=1}^{B} \overline{R}^2_{(i)} \; x \; \Delta_i}{\sum_{i=1}^{B} \overline{R}^2_{(i)}} \tag{7.17}$$

7.3.3 The corrected Akaike Information Criteria (AICc)

The $AICc$ approach orders the model according to their performance, AIC and the weight of each of them. The AIC is calculated as follows:

$$AIC = -2log(\delta(M/Data)) + 2K \tag{7.18}$$

Where $\delta(M/Data)$ represents the likelihood of the model M given the data, and K is the number of parameters in the model. In case the errors are normally distributed, the AIC is calculated as follows:

$$AIC \;\approx\; n \; log(\frac{\sum_{i}^{n} \widehat{\in}^2_i}{n}) \;+\; 2K \tag{7.19}$$

Where $\widehat{\in}^2_{.}$ denotes the residual sum of squares of the model and n the sample size. The AIC estimate is prone to bias if the sample size is small with many predictors, which is the case in this study. It is therefore recommended to proceed by the corrected version of the $AICs$ (Burnham and Anderson, 2014), estimated as follows:

$$AIC_c = \; AIC \; + \; \frac{2K(K+1)}{n \; - \; k \; - \; 1} \tag{7.20}$$

The $AICc$ of each candidate model is weighted as follows:

$$W_i \;=\; exp\left(-\frac{1}{2}(AICc_i \; - \; AICc_{min})\right) \tag{7.21}$$

Where AICc_{min} represents the smallest value of AIC_c among the values for the B candidate models. The quantity of interest can be expressed as follows:

$$E(\Delta) = \frac{\sum_{i=1}^{B} W\ x\ \Delta_i}{\sum_{i=1}^{B} W_i} \tag{7.22}$$

7.3.4 Bayesian information criteria (BIC)

The BIC, also referred to as the Schwartz's information criteria weight models by assuming that they all have an equal prior probability, is expressed in the following formula:

$$BIC\ =\ -2log(\delta(M/Data)) + 2K\ x\ log(n) \tag{7.23}$$

When the errors of all models are normally distributed with a constant variable, the BIC can be estimated as such:

$$BIC\ \approx\ n\ log(\frac{\sum_i^n \widehat{\in}_i^2}{n})\ + K\ \times\ log(n) \tag{7.24}$$

The weight of each model corresponds to the joint probability $P(M_i|D)$ as expressed in the following formula:

$$W_i\ =\ P(M_i|D)\ =\ exp\left(-\frac{1}{2}\ (BIC_i\ -\ BIC_{min})\right) \tag{7.25}$$

The posterior probability of each model is therefore the rapport of the joint probability over the marginal probability.

7.3.5 Bayesian moving average (BMA)

All Bayesian approaches are based on the prior beliefs of the candidate models. For the case of BMA, all models, as well as their corresponding parameters, receive a uniform prior distribution. Therefore, the posterior distribution will be estimated as follows:

$$P(M|D)\ =\ \frac{1}{B} \times \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}}\ exp\left\{-\ \frac{\widehat{\epsilon}_i^2}{2\sigma^2}\right\} \tag{7.26}$$

Where $1/B$ stands for the prior probability on each model M_i, and the rest of the equation represents the likelihood of the considered model given the observed data.

7.4 Prediction of the maize yield

The first phase of the model is described in equation 7.1, where the observed climate predictors from CRU (large scales) and ERA5 (local scales) are used to calibrate the model. The second phase consisted of transferring the developed model to the REMO for predicting the future yield and yield changes. As for the observed predictors, the REMO predictors were also standardised. The mean and standard deviation normalization were calculated for each predictor using the REMO mean and standardization covering the period from 1982 to 2012. The following equation indicates how to obtain the predicted yields, as both the maize yield and the predictors were standardized.

$$Y_i = (Y_i' \times S_{n-1,Y}) + \overline{Y} \tag{7.27}$$

In this equation, Y_i denotes the yield data in t/ha and Y_i' the yield estimate from the model. $\overline{Y}$ and $S_{n-1,\,Y}$ represent respectively the mean and the standard deviation of the observed maize yield relative to the 1982–2012 period. The yield changes for the inter-comparison were calculated relative to the 1982–2012 period. It is therefore mathematically expressed as such:

$$\Delta Y_i = \frac{Y_i - \overline{Y}}{\overline{Y}} \times 100 \tag{7.28}$$

Where ΔY_i denotes the yield changes relative to the inter-comparison period (1982-2012) and $\overline{Y}(1982/2012)$, the average of the observed yield.

8. Climate Impact Assessment

This chapter is divided into two sections. The first section deals with the statistical modelling of the maize yield from model development to future yield changes. The second section assesses the impact of climate change directly on WRSI.

8.1 Statistical crop model for Maize in tropical Africa

This section first describes the group of predictors, as well as their association with the maize yield. After this, the performance of the statistical model in predicting the present-day and future maize yield changes in tropical Africa is evaluated.

8.1.1 Traditional and process-based Local Predictors

Traditional predictors refer to the most commonly used climate variables used to assess the impact of climate change on crop yields. Those predictors are the mean temperature, the maximum temperature and the minimum temperature, as well as the mean precipitation. These variables were all derived from ERA5 datasets. In contrast, the process-based predictors are those derived from the WRSI, calculated in chapter 6. This concerns the global WRSI (WRSI-glo), the WRSI of the initial crop stage (WRSI-int), the WRSI of the crop development stage (WRSI-dev), the WRSI of the mid-season (WRSI-med) and the WRSI of the late season (WRSI-late). These nine variables are called locals because they are used at the grid point level.

Figure 8.1 shows the means of the WRSI-glo over five years, as well as the linear trend calculated from 1979 to 2018. The results indicate that WRSI-glo is very high – exceeding 90 % – in the Guinea coasts: south of the Gambia, Guinee Bissau, Sierra Leonne, Guinea and the south of Mali. Another pattern of high WRSI-glo is reported in the centre of

Nigeria. This implies that the water availability for maize production is suitable in these regions. The WRSI-glo is around 80 %, particularly in the Republic Democratic of Congo, Republic of Congo, Cameroon, Gabon, Central Africa, the south of Sudan and the nord of Angola. However, in the last two decades (1999–2018), in the south of Cuvette Centrale, precisely in the province of Bandundu of the Republic Democratic of the Congo, the WRSI-glo has substantially decreased. In arid and semi-arid regions, such as the Sahel, the desert of Namibia and Somalia, the WRSI-glo is extremely low, below 10 %. This implies that maize could grown adequately without irrigation. The linear trend indicates that tropical Africa is experiencing a decrease in maize yield performance in general. A significant trend is observed at 10° north, starting from West Africa (Senegal, southern Mali, Burkina Faso, Ivory Coast, Nigeria, Cameroon, Central Republic of Africa and South Soudan). Another significant negative trend of WRSI is observed in the centre of the Republic Democratic of the Congo; in east Africa between Rwanda, Burundi, Uganda and the North of Tanzania; and the east of Ethiopia. From 1979 to 2018, the Central African Republic lost 22 % of maize yield performance on average, 9 % on average was lost in the Republic Democratic of Congo, 11 % was lost in Benin, and 3 % was lost in Nigeria. However, some regions of Africa are experiencing an increase in global WRSI. In Angola, for example, the WRSI-glo has increased by 6 % (Table 8.1). Details about the percentage of the WRSI-glo per country are presented in Table 8.1.

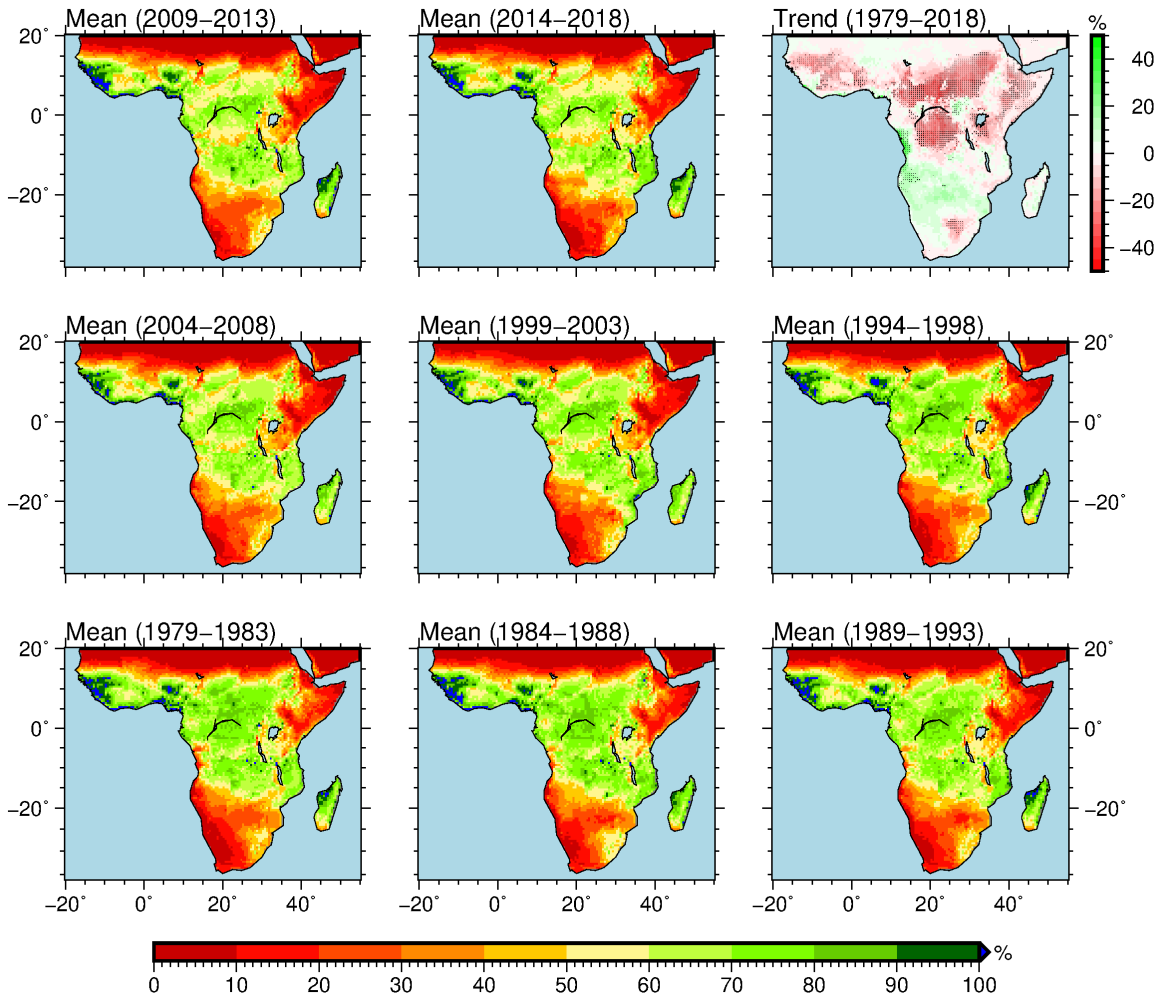

Figure 8.1: Spatial distribution of the five-year means of WRSI-glo and linear trends (extreme top-right).

The WRSIs calculated for each stage of the maize growing season are presented in Figure 8.2 The initial stage (Figure 8.2a) is characterised by high values of WRSI-int, exceeding 100 % in the Congo Basin. During this stage, the maize water requirement is low (with the crop coefficient being close to 0.4), but the precipitation is abundant due to the permanent presence of the ITCZ. In addition, the initial stage starts in June/July – which is the peak of precipitation – in the Congo Basin. The value of the WRSI-dev in the developing period is between 70 and 100 % in most regions of tropical Africa, except in the arid regions (Figure 8.2b). At this stage, the crop needs almost double the water required during the initial stage. The peak of water requirement for maize is reached in the mid-season, where the crop becomes mature (the crop coefficient is about 1.2) (Figure

Table 8.1: Mean, maximum and minimum values of WRSI in selected African countries

Countries	Min (%)	Max (%)	Mean (%)
Angola	-22.36	-22.36	-22.36
Benin	-14.97	-8.22	-11.73
Congo	-10.03	24.11	0.04
Democratic Republic of the Congo	-29.47	18.89	-9.88
Burundi	-5.27	-4.44	-4.86
Cameroon	-25.37	2.83	-12.99
Central African Republic	-31.97	-7.44	-22.7
Equatorial Guinea	-1.34	-0.84	-1.09
Gabon	-6.25	4.83	-1.41
Ghana	-13.49	16.39	-3.72
Guinea	-17	9.48	-3.75
Kenya	-18.5	0.83	-7.35
Madagascar	-3.34	10.79	1.73
Mali	-23.43	1.56	-6.34
Malawi	-9.87	10.69	-3.19
Guinea-Bissau	-11.53	18.78	-0.52
Senegal	-18.69	2.81	-9.14
Sudan	-25.06	3.76	-9.71
United Republic of Tanzania	-21.83	7.83	-6.77
Namibia	1.77	12.9	7.86
Zimbabwe	-1.03	13.66	6.62

8.2c). Consequently, the WRSI-mid is low in the humid region, between 50 and 65 %. The water needed decreases somewhat in the late season, where the crop coefficient is around 0.7, but the value of WRSI-late in the Congo Basin is above 90 %, exceeding even 100 % in some regions (Figure 8.2d). High values of WRSI-late are also observed in West Africa on the Guinea coast and in Burkina, Benin and Ghana.

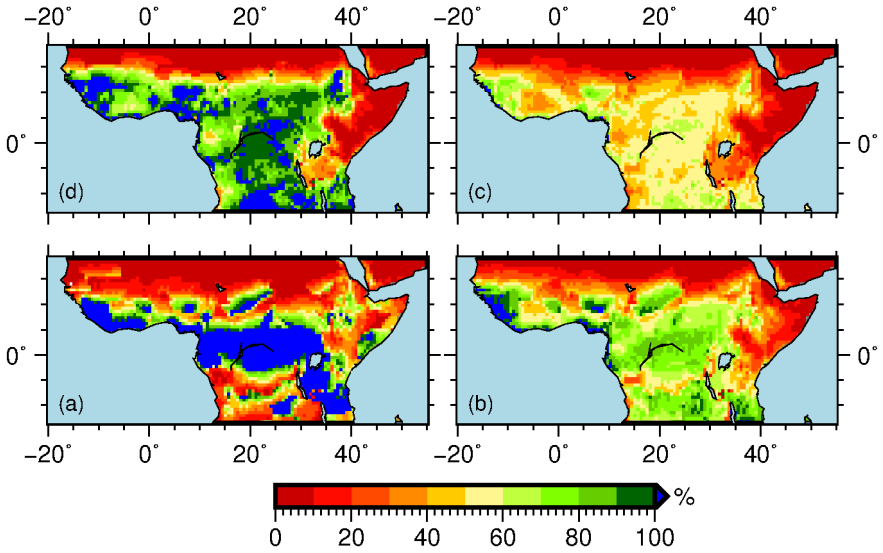

Figure 8.2: Spatial distribution of the WRSI calculated during different maize growing season stages: (a) initial; (b) developing; (c) mid-season; and (d) late season.

Figure 8.3 presents the linear correlation between each local predictors (processed-based and simple) with crop yield. It is important to note that neither maize yield nor the predictors were detrended. The results show that the correlation pattern between the maize yield and the three temperature predictors (maximum, minimum and mean) are almost similar (Figure 8.3a, 8.3b, 8.3d). Significant negative correlations are found in eastern Africa, specifically Kenya, Uganda and Tanzania. This is also true for north-western Africa and southern Africa. However, the three temperature predictors are positively correlated with maize yield in some humid regions (Nigeria and the Republic Democratic of the Congo) and the north of east Africa (Ethiopia and North of Kenya). The correlation pattern between the maize yield and the five water requirement satisfaction index (WRSI-glo, WRSI-int, WRSI-dev, WRSI-mid and WRSI-late) is also almost identical (Figure 8.3e, 8.3f, 8.3g, 8.3h, 8.3i). The maize yield is positively correlated with the five WRSIs in east Africa and in the north of western Africa (Sahel regions). The correlation

between crop yield and WRSIs in equatorial Africa, specifically in the Democratic Republic of the Congo, is negative. The spatial pattern of the mean precipitation is close to one of the five maize WRSIs (Figure 8.3c).

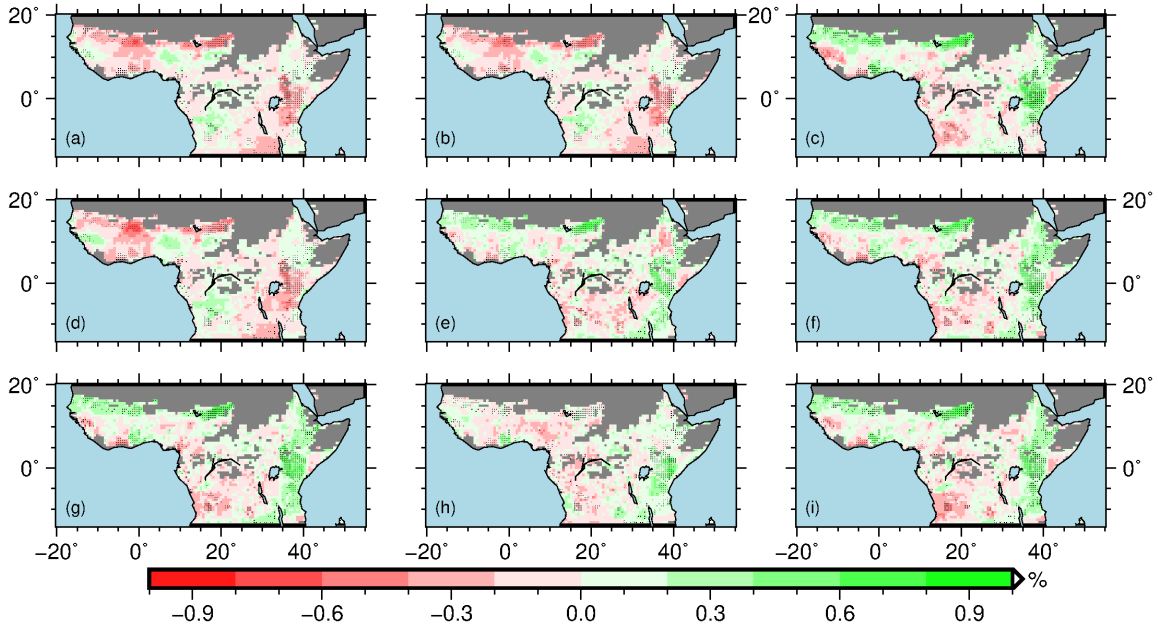

Figure 8.3: Linear correlation between maize yield and local climate variables: (a) maximum temperature, (b) minimum temperature, (c) mean precipitation, (d) mean temperature, (e) WRSI-lat, (f) WRSI-mid, (g) WRSI-glo, (h) WRSI-int, (i) WRSI-dev. Black dots indicate places where correlations are significant at the 5 % significance level.

8.1.2 Large-scale climate predictors

Large-scale climate predictors are derived from EOFs, as explained in Subsection 5.2.5. The first EOF was applied to the precipitation and temperature grid points, separately. The second EOFs – from which came the final large-scale predictors – were applied to the precipitation and the temperature PCs obtained from the first EOF.

Figure 8.4 presents the leading EOFs and PCs of September–October–November precipitation in central Africa. The plotted PCs were normalised using standard deviation based on the time period from 1960 to 1990 and filtered using nine years moving average. The first EOF pattern explains 13.8 % of the total precipitation variance, while the second explains 11.2 %. The first EOF pattern is homogeneous with low values in Sahara and

Kalahari. These values increase equatorward, and the maximum is reached in east Africa and in the west of the Congo Basin, along the Atlantic coast. Its associated principal component, PC1, exhibits a multi-decadal variability with a periodicity of almost 30 years. Between 1908 and 1919, the PC1 was wetter than the baseline period. Other wet periods are observed from 1929 to 1947 and from 1971 to 2000. The periods 1950–1980 and 2000–2018 are considered drier, but the curve indicates a slight recovery. After projecting REMO precipitation anomalies (simulated under RCP8.5 and RCP4.5 scenarios) into the EOF spatial pattern of the CRU dataset, we obtained PC1-REMO (RCP4.5) and PC1-REMO (RCP8.5). Compared to PC1, the present-day PC1-REMO (RCP8.5) reproduces the temporal variability of PC1 quite well and seems to be coherent in projecting future variability. PC1-REMO (PCP85) captures the drier conditions after 2000 well and projects that the periods between 2022–2054 and 1968–2100 will be wetter than the baseline period. PC1-REMO (RCP4.5) poorly captures the drying condition after 2000, but represents the amplitude of the PC1 variability better than PC1-REMO (PCP85). The direction of both PC1-REMO is opposite after 2080. PC1-REMO (RCP4.5) projected a drier period after 2080, while PC1-REMO (RCP8.5) indicates a wetter period. In contrast to the EOF1 spatial pattern, EOF2 presents a dipole spatial pattern of the opposite sign. Positive precipitation values are found in West Africa, and negative values are found in East Africa. Its associated PC displays a higher amplitude and longer time period. Therefore, the period 1925–1980 was wetter than the baseline period. Wetter periods are also observed from 1995 to 2018. The corresponding PC2-REMO (RCP8.5) and PC2 (REMO) RCP4.5 do not reproduce the temporal variability of PC2, but both indicate, as per PC2, that the years after 1995 were wetter than the baseline period.

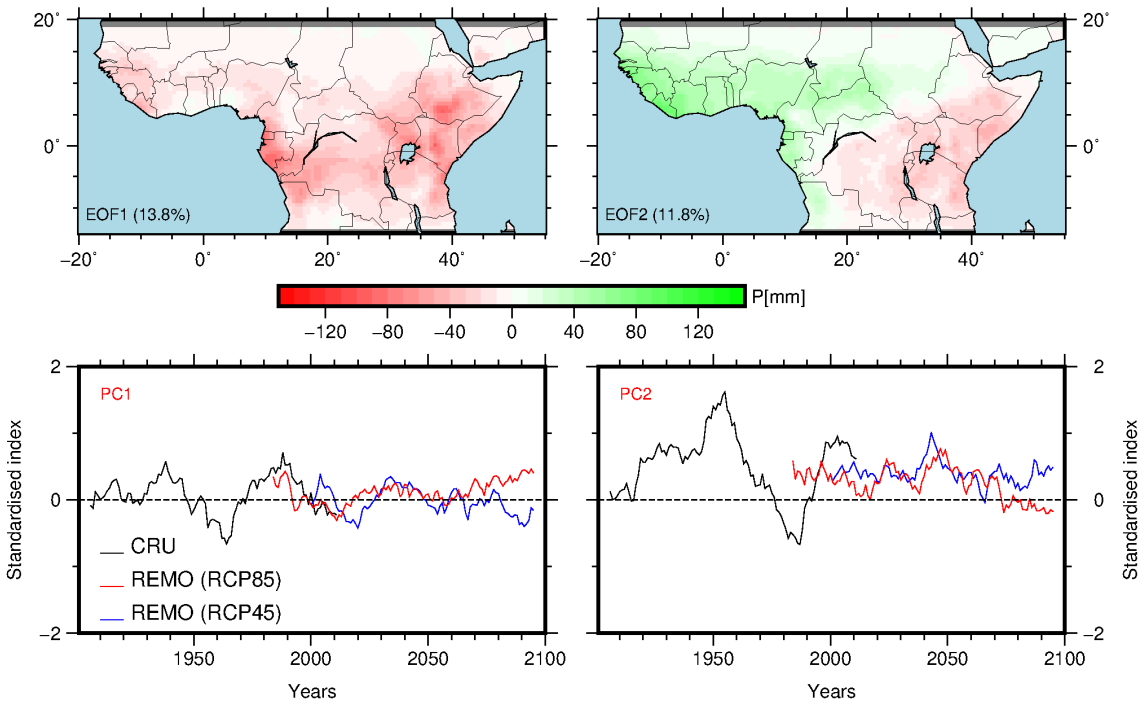

Figure 8.4: Nine years moving average time series of the two leading principal component for September-October-November precipitation sums of CRU and REMO datasets, and their associated spatial pattern. PCs have been standardised relative to 1961–1991 climatology mean.

Figure 8.5 shows the two-leading precipitation EOFs of the September–October–November season and their associated PCs. In contrast to the precipitation, the temperature dataset was not detrended to better interpret the curve, but we used the detrended version in the statistical model. The spatial pattern indicates that EOF1 explains 49.2 % of the total temperature variance and presents a homogeneous spatial pattern with high values of eigenvectors in east Africa and north-east Africa. Its associated PC1 indicates that, from 1901 to 1925, the temperature was colder than the baseline average (1960–1990). Another colder period was observed between 1945 and 1978. The temperature starts to warm from 1979 up to 2018. Both PC1-REMO (RCP4.5) and PC1-REMO (RCP8.5) simulated the present-day temperature very well. Concerning the future simulation, PC1 REMO (RCP8.5) captures very well the linear trend of observed temperature while PC1-REMO (RCP4.5) fails to reproduce the observed linear trend. The EOF2 pattern displays a dipole spatial pattern of opposite signs between North Africa and East Africa. The associated PC2 time series indicates that between 1901 to 1948, the temperature was warmer

than the average baseline. From 1955, the temperature recovered and became warmer again in 1980. The REMO models do not capture the present-day simulation.

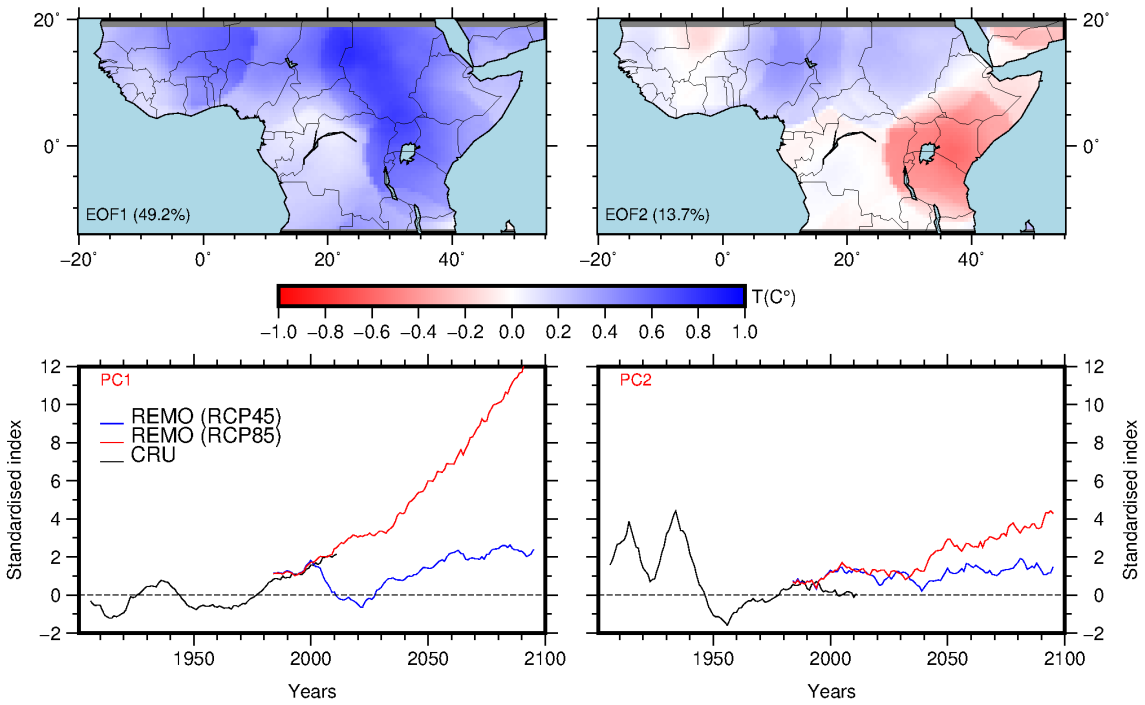

Figure 8.5: Nine years moving average times series of the two leading principal components for September-October-November temperature mean of CRU and REMO non-detrended datasets, and their associated spatial pattern. PCs have been standardised relative to 1961–1991 climatology.

Figure 8.4 and Figure 8.5 are only examples of the EOF analysis computed for the period September–October–November, but the gridded-based PCs, which were further used in the second EOF analysis, were computed using the monthly dataset, as illustrated in figure 5.5, and were all detrended. The PCs derived from the second EOF analysis were correlated with climate indices in order to select, as final predictors, variables that have physical meaning (significant correlation with climate indices). Table 8.1 displays the correlation between PCs and climate indices. It can be observed that the four leading PCs are correlated with climate indices and explain 55 % of the total data variability. They are therefore selected as large-scale climate variables for modelling maize yield. The remaining PCs are considered as simple mathematical constraints because they do not convey any physical insight. In detail, PC1 is significantly correlated with the TNA (r = 0.56, p = 0.01) and the NAO (r = - 0.46, $p < 0.01$), TSA (r= 0.39, $pvalue$ = 0.05), the

Atlantic SST (r = 0.46, p = 0.05) and the Indian ocean SST (r = 0.41, p = 0.05). PC2 is highly correlated with the nino3 index (r = 0.72, $p < 0.01$), MEI (r = 0.76, $p < 0.01$), SAOD (r = -0.68, $p < 0.01$), TSA (r = - 0.52, $p < 0.01$), Indian ocean SST (r = 0.57, $p < 0.01$), the TNA (r = 0.4, p = 0.05) and the Atlantic SST (r = 0.36, $p < 0.01$). PC3 is significantly correlated with Atlantic SST (r = 0.68, $p < 0.01$), the TSA (r = 0.59, $p < 0.01$), the SAOD (r = 0.47, $p < 0.01$), the NAO (r = -0. 43, p$value < 0.01$) and the IMI (r = - 0.38,$p < 0.01$). PC4 is significantly correlated with the QBO (r= 0.56, $p < 0.01$), the IMI (r = - 0.56, $p < 0.01$) and IOD (r = 0.43, p = 0.05).

Table 8.2: Linear correlation between climate indices (Atlan_SST = Atlantic Seas Surface Temperature; Indi_SST = Indian Seas Surface Temperature; TNA = Tropical North Atlantic; TSA = Tropical South Atlantic; MEI = Multivariate ENSO Index; NAO = North Atlantic Oscillation; SAOD = South Atlantic Ocean Dipole; QBO = Quasi-Biannual Oscillation; IMI = Indian Monsoon Index; DMI = Indian Monsoon Dipole) and maize yield. * denotes the corrolation is significant at 5 5% significant level, and ** at 1 % significant level.

Var.	Nino3	SAOD	QBO	DMI	IMI	NAO	MEI	SAMI	TSA	TNA	Atlan_sst	Ind_sst
Pc1	0.02	0.12	-0.31	-0.1	0.03	-0.46**	0.13	0.26	0.39*	0.56**	0.46*	0.41*
Pc2	0.72**	-0.68**	-0.03	0.09	-0.03	0.24	0.76**	-0.32	-0.52**	0.4*	-0.36*	0.57**
Pc3	-0.15	0.47**	0.07	-0.27	-0.38*	-0.43*	-0.19	0.26	0.59**	0.35	0.68**	0
Pc4	0.35	0.03	0.56**	0.43*	-0.56**	0	0.15	0.09	0.02	0.09	0	0.24
Pc5	-0.18	-0.02	0.33	-0.21	0.09	0.42*	-0.21	0.28	-0.04	-0.03	-0.04	-0.01
Pc6	0.08	0.07	-0.13	0	0	-0.06	0.03	-0.03	-0.08	-0.11	-0.12	-0.36
Pc7	0.11	-0.18	-0.14	-0.09	0.16	-0.02	0.09	-0.12	-0.1	-0.19	-0.16	-0.17
Pc8	0.26	-0.1	0.22	0.1	-0.33	-0.09	0.18	-0.02	0.04	0.03	-0.05	-0.03
Pc9	-0.17	-0.08	-0.22	0.14	-0.02	0.19	-0.16	-0.28	0.1	-0.29	0.02	0.13
Pc10	0.13	-0.01	0.08	0.2	0.17	0.03	0.12	0.33	0.03	0.22	0.08	0
Pc11	0	0.05	-0.12	-0.05	0.05	0.18	-0.02	-0.18	0.02	0.15	0.02	-0.17
Pc12	-0.24	-0.19	0.13	-0.51**	0.2	-0.05	-0.19	-0.2	-0.2	0.12	-0.04	-0.14
Pc13	-0.03	-0.06	-0.19	-0.01	-0.07	-0.07	-0.1	-0.09	0.09	-0.02	0.07	0.09
Pc14	0.06	-0.12	0	0.24	0.15	-0.24	-0.05	0.38*	-0.09	0.25	-0.06	0.11
Pc15	-0.01	-0.1	-0.08	0.07	0.16	0.05	0.05	-0.09	-0.04	-0.06	-0.03	-0.16
Pc16	-0.18	0.15	0.13	-0.15	0.17	-0.05	-0.15	-0.17	0.1	0.13	0.09	-0.3

8.2 Evaluation of present-day crop models

This section presents and evaluates cross-validated models according to the group of predictors. It also assesses the impact of the multi-model averaging technique on the maize yield model.

8.2.1 Impact of the group of predictors on the Maize yield model

According to the nature of predictors, we set four models: (a) a model developed with all predictors (**M1**); (b) a model with local process-based predictors (**M2**); (c) a model with local traditional predictors (**M3**); and (d) a model with large-scale predictors (**M4**). The explained variables of these models are computed at the $RMSEcv$ equal to the minimum values, as explained in Section 7.2.

Figure 8.6 presents the spatial distribution of the explained variance of the model developed with different groups of predictors and their associated empirical probability density function. The results indicate that the impact of all predictors on maize yield form a complex pattern and vary across different geographical locations (Figure 8.6a). The global model (**M1**) explains almost 28 % of the yield variability, and its empirical probability density curve reaches the maximum value of 27 % (Figure 8.6e). Obviously, the **M1** model outperforms the other models (**M2**, **M3**, **M4**). **M1** is particularly skilful and consistent in modelling maize yield in the western parts of the region, where the average explained variance of the model is around 51 %, reaching 60 % in some regions such as the south of the Ivory Coast and Ghana. **M1** seems also to be performant in Kenya. The spatial distribution of the explained variance of **M4** (Figure 8.6d) is quite close to that of **M1**, except with regard to Tanzania, where **M4** performed poorly. The empirical probability density curve of **M4** is also similar to that of **M1**.The main difference between the models is that the empirical probability density curve of **M4** indicates a higher density of explained variance ranging from 0 % to 15 %, compared to **M1** with a lower density of explained variance from 16 % to 70 %. Compared to **M1** and **M4**, the models developed using local predictors (**M2**, **M3**) are much less performant in modelling maize yield, but in some regions, such as Kenya, they appear to be skilful. Based on the spatial distribution, **M2** and **M3** do not differ across the study area. However, the empirical probability distribution curve indicates that **M2** is better than **M3**.

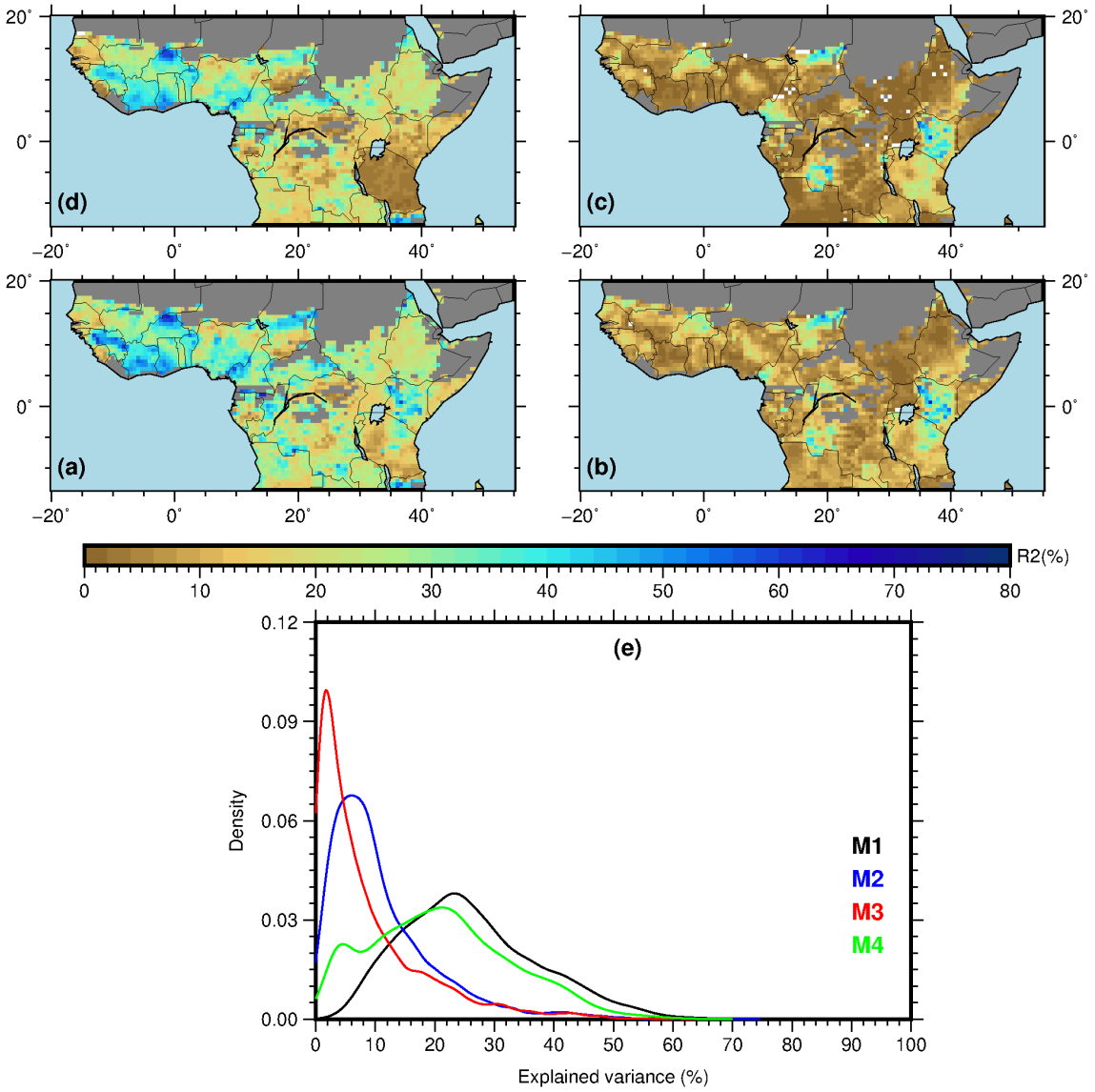

Figure 8.6: Mean explained variance in % of the cross-validated regression at $RMSE_{CV}$ = Minimum. (a) Model with all predictors, (b) model with simple local predictors, (c) model with WRSI predictors, (d) model with large scale-predictors, and (e) the empirical probability density function of the explained variance (**M1** = model with all predictors; **M2** = model with WRSI predictors; **M3** = model with traditional predictors; **M4** = model with large-scale predictors).

8.2.2 Impact of multi-model averaging techniques on the Maize yield model

The multi-model averaging techniques are applied to the model with all predictors presented in Section 7.3. We therefore used BMA, AIC, AICc, BIC, adjusted R-squared and mean. The explained variance, the RSME and the empirical probability density curve are used to rank the performance of each multi-model averaging technique. The impacts of each of these six multi-model techniques on maize yield are displayed in Figure 8.7. Their associated empirical probability density curves are presented in Figure 8.8. The results indicate that multi-model averaging techniques based on BMA, BIC, AIC and AICs performed very well and their spatial patterns are very similar (Figure 8.8). However, the empirical probability density curve indicates that BMA is slightly better than the other models. Therefore, the model (**M1**) based on BMA as a multi-model averaging technique predicts future maize yield changes.

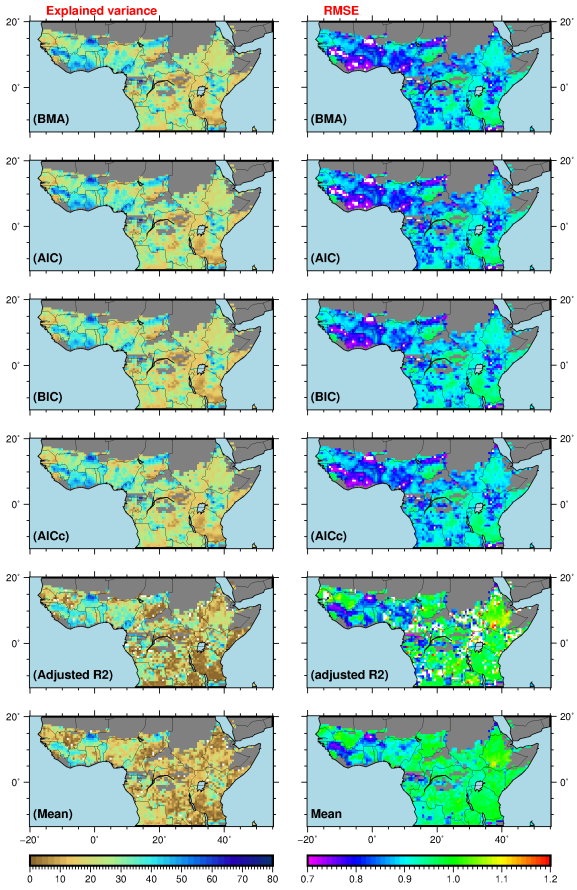

Figure 8.7: Spatial distribution of the explained variance and the root mean squared errors of different multi-model averaging techniques (BMA = Bayesian Moving Averaging; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; AICc = corrected Bayesian Information criterion).

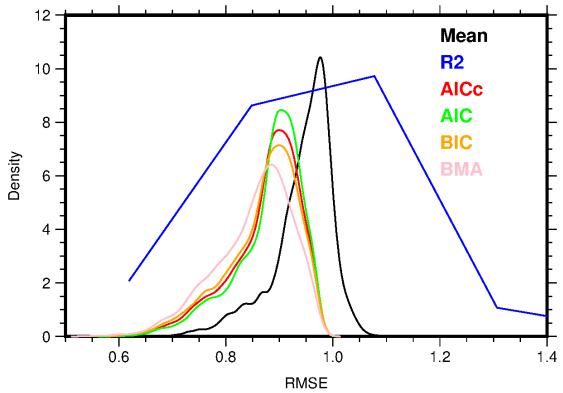

Figure 8.8: Empirical probability density function of the root mean squared error for different multi-model averaging techniques.

8.2.3 Model variable selection

As described in Section 7.2, not all input predictors are used in the model. In order to ensure physical relevance between predictors and predictand, the model selects the optimal number of predictors through a crossed-validated mechanism. Therefore, the optimal number of predictors is obtained when the MSE_{CV} reaches the first relative minimum. Figure 8.9 presents the spatial distribution of the optimal number of predictors selected and used in the regression. The results indicate that the optimal number of predictors forms a complex and heterogeneous pattern. In general, the first relative minimum in each grid point is reached when one or four predictors are selected, but on average, two robust predictors are selected per grid point.

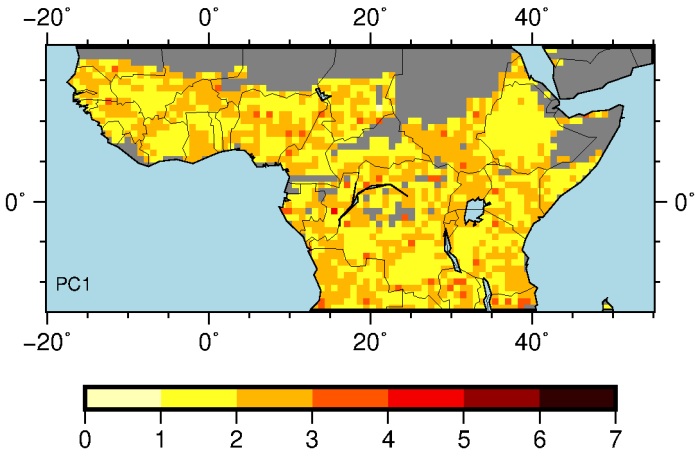

Figure 8.9: Mean number of cross-validated predictors at RMEcv = minimum.

Another important aspect of this model is how often a given predictor is selected in every grid point over 100 iterations (Figure 8.10). The results suggested that large-scale predictors are more selected than local ones. Among large-scale predictors, PC2 – which is strongly associated with ENSO and weakly associated with the Indian SST – is preferably selected, being selected 36 times on average for the whole region. In addition, it seems to be the dominant factor in the western parts of the study area, across the Guinean coast and in the southwestern parts of the region. The following most important predictors are PC4 (selected 26 times on average) and PC1 (selected 23 times on average). PC4 is highly associated with QBO and PC1 with the TNA. The impacts of PC4 on the maize yield model are very high in the eastern parts of the study area, particularly in

the highlands of Ethiopia, where PCs are selected more than 80 times, on average. The results also indicate that the local predictors influence maize yield more in colder regions, such as in the eastern parts of the study area. Among the nine local predictors, the mean temperature is the most important one because it was selected 13 times on average. The second most important local predictors are the WRSI-dev and the precipitation.

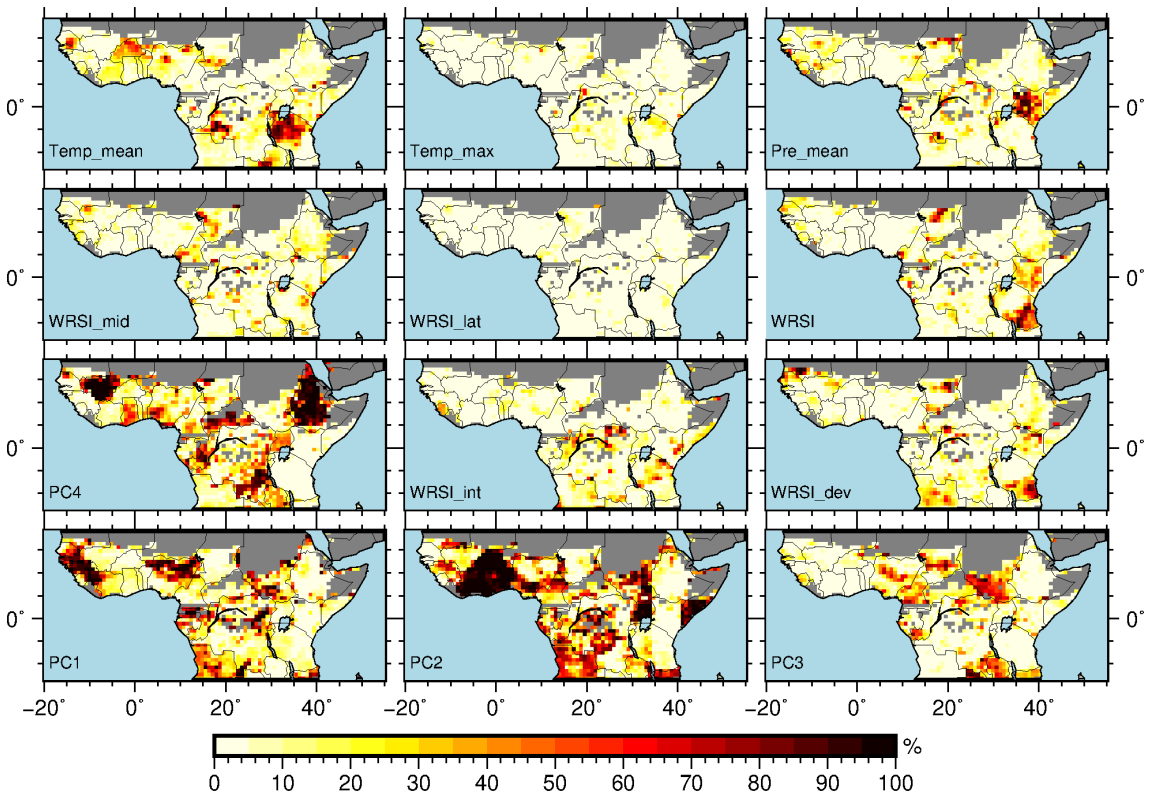

Figure 8.10: Number of times a given predictor is selected in the cross-validated regression, over 100 interations.

Apart from ranking the predictors in order of importance, the model also counts the number of times, over 100 iterations, a coefficient of a given predictor is positive or negative. Figure 8.11 presents a spatial distribution of how many times a positive coefficient is selected per predictor in each grid points. The results suggest that PC2, which is the most dominant predictor in the study region, positively influences maize yield in the Congo Basin, especially in the Democratic Republic of the Congo. The relationship between PC2 and maize yield is also positive in the south of Somalia. The PC1 coefficient is generally positive and more frequent in the extreme western parts, in northern Nigeria and

in Angola. PC4 is positively correlated with the maize yield in the Ethiopian highlands. The spatial distribution of the positive coefficients for the nine local predictors is also similar with that of the negative coefficients, as displayed in Figure 8.12.

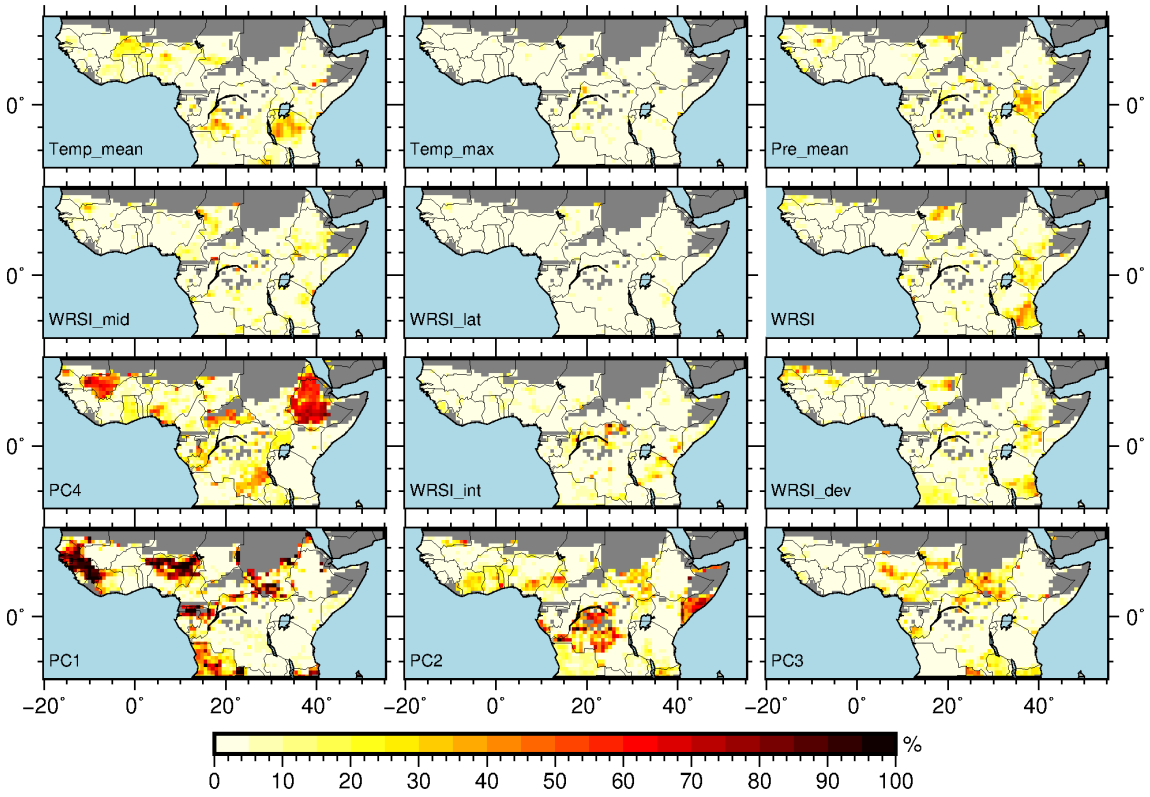

Figure 8.11: Number of times a given predictor yields a positive coefficient in the regression model, over 100 iterations.

Figure 8.12 presents the spatial distribution of how many times a negative coefficient is selected per predictor in each grid points. The results indicate that PC2 negatively influences maize yield in western Africa. Negative coefficients are not frequently selected with PC1. The spatial pattern of negative coefficients in PC4 is similar to that of positive coefficients, except in the highlands of Ethiopia, where positive coefficients are preferred.

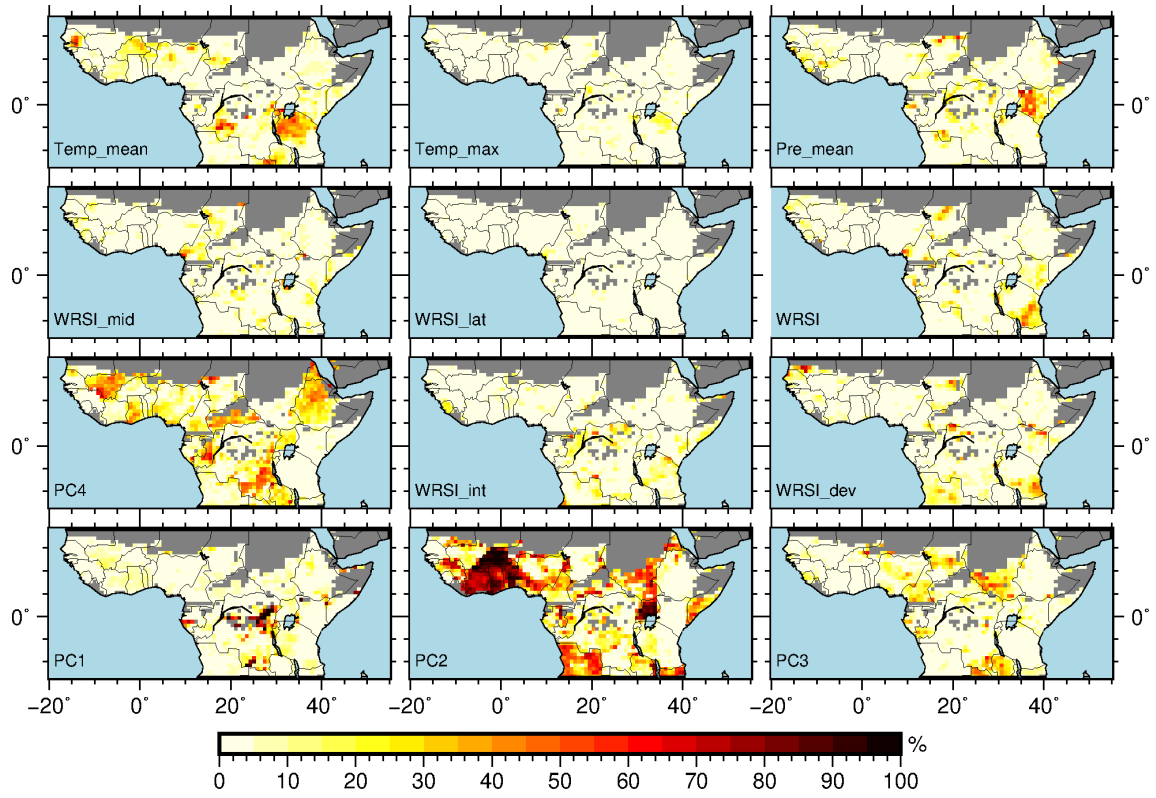

Figure 8.12: Frequency of each predictor yielding a negative coefficient in the regression model over 100 iterations.

8.3 Modelling Maize yield under the future climate change Condition

This section presents the observed present-day and future climate predictors selected to model the maize yield. These variables are used in the second phase of the regression model to forecast the future yield change of maize under different climate emission scenarios. The impact of different weighing techniques on maize projection is assessed, as is

the contribution of different groups of climate predictors on future maize yield changes.

8.3.1 Projection of maize yield changes under climate change conditions

The projection of maize yield assumes that the statistical transfer function used to develop the maize yield remains stationary in the future. The REMO-simulated predictors are put into the cross-validated equation to model future maize yield changes. Figure 8.13 presents present-day and future projections of the median maize yield change under the RCP8.5 and RCP4.5 emission scenarios, as well as their associated confidence intervals based on 100 bootstrap samples. The projection is based on the BMA multi-model technique and the changes in each grid are relative to the baseline period (1982–2016). The results indicate that, despite the climate scenario being used, there is a decrease in maize yield over tropical Africa. Under the RCP8.5 emission scenario, the crop yield decreases in 2018 and peaks in 2060 with 20 % yield reduction. It starts to recover from 2061 until the end of the century but is still negative compared to the baseline period. Under the RCP4.5 emission scenario, the projected maize yield is on average reduced by 5 % from 2050 to the end of the century.

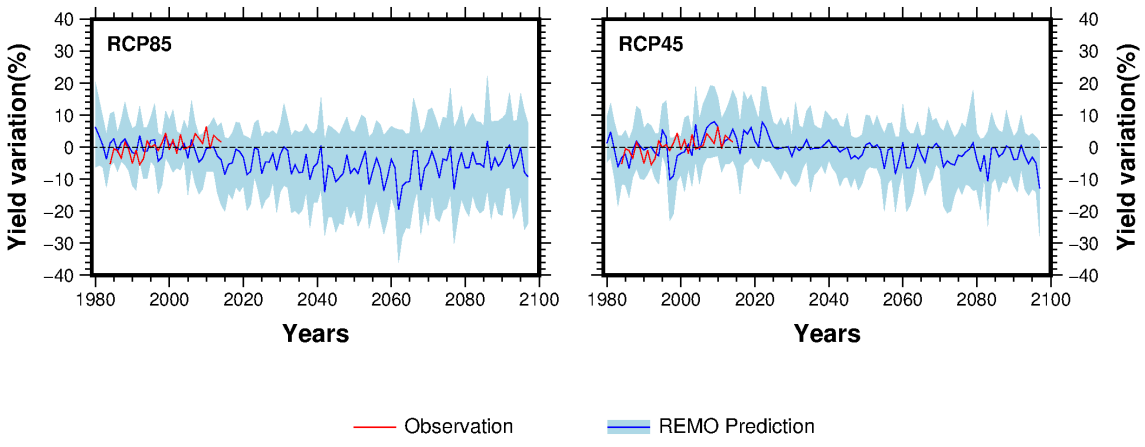

Figure 8.13: Projection of median maize yield changes under RCP8.5 and RCP4.5 climate scenarios and their corresponding confidential intervals at the 5 % significance level. The yield changes are calculated based on average maize yield production from 1982 to 2016.

Table 8.3: Nonlinear trend detection on the projected maize yield using Mann Kendall test

	RCP8.5				**RCP4.5**			
	1982 - 2100	1982 - 2018	2018 - 2060	2061 - 2100	1982 - 2100	1982 - 2018	2018 - 2060	2061 - 2100
Z	-5.38	-3.19	-1.67	1.58	-4.2	2.7	-2.72	-0.3
S	-2317	-265	-150	122	-1811	229	-244	-24
p	0.001	0.001	0.09	0.113	0.001	0.005	0.006	0.7

Table 8.2 reports the results of the non-parametric trend analysis of global maize changes time series using a Mann–Kendall test and therefore completes the information conveyed in Figure 8.13. Details about the test are provided in Subsection 5.2.4. The results show that, in the RCP8.5 emission scenario, crop yield changes significantly decrease during the entire period (1982–2100), particularly from 1982–2018, at the 1 % significance level. Between 2018 and 2060, a statistical decrease is also observed, but at the 5 % significance level. The median yield changes increase from 2061 to 2100, but not significantly ($p = 0.113$). The median maize changes times series do not show a clear trend under RCP4.5; a statistically significant decrease is observed for the entire time series at the 1 % significance level. From 1982 to 2018, the trend increases statistically at the 5 % significance level. From 2018 to 2060, the trend decreases at the 5 % significance level; it still decreases from 2060 to 2100, but not significantly. The sequential Mann–Kendall test statistics for abrupt change detection indicate that the years 2013 and 2042 are considered as turning points for RCP8.5 and RCP8.5, respectively.

Figure 8.14 provides more detail on the median maize yield changes by separately analysing the trend across central, western and eastern tropical Africa. This trend is interpreted with the results of the Mann Kendal test. The results show that the median maize yield changes are projected to decrease by the end of the century under RCP4.5 and RCP8.5 emission scenarios. The Mann Kendal test statistics are significant (p <0.001) for both emission scenarios. The sequential Mann Kendal test shows that the abrupt year changes under RCP8.5 and RCP4.5 are 2065 and 2045, respectively. In eastern Africa, the future median maize yield changes are expected to increase significantly under RCP8.5 by the end of the century (p <0.001), with an abrupt change in 2055. However, a non-significant decrease in median yield change is observed under RCP4.5 (p < 0.06). In central Africa, the maize yield variability fluctuates around the baseline mean under RCP4.5 and RCP8.5.

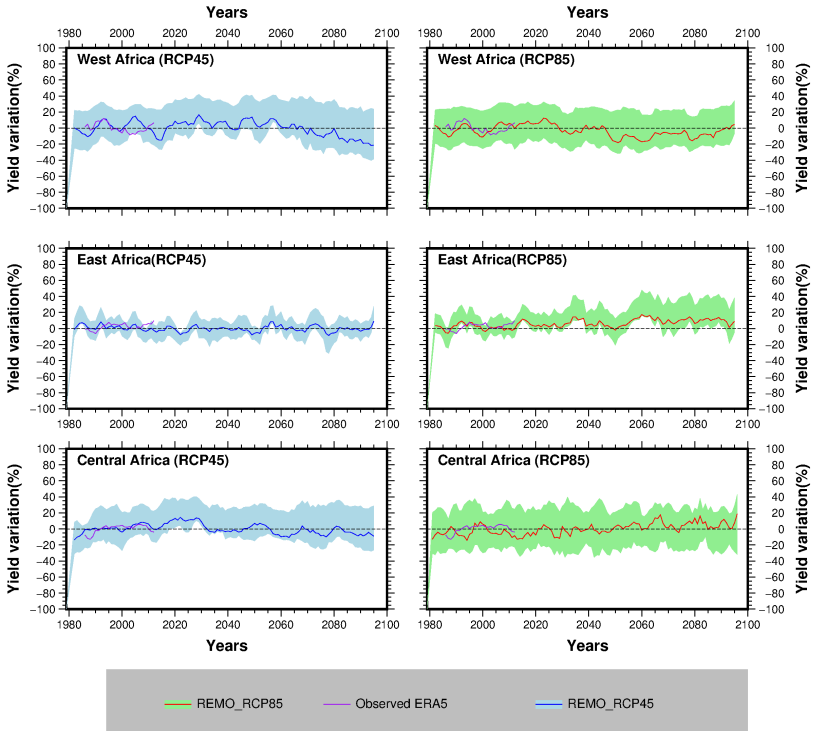

Figure 8.14: Projection of median maize yield changes under RCP8.5 and RCP4.5 climate scenarios and their corresponding confidential intervals at the 5 % significance level for the central, eastern and western parts of the study area. The yield changes are calculated based on average maize yield production between 1982 and 2016.

The median maize yield change pattern is presented in Figure 8.15. The lower and upper confidential intervals at the 5 % significance level, indicating the projection uncertainties, are also presented. The results indicate that the impacts of all predictors on future median maize yield vary according to location and climate emission scenario. Under RCP4.5, the spatial median maize yield changes have not substantially changed (very close to zero), but a slight decrease in future median maize yield change over the entire study areas is observed, especially at the end of the century. A reduction in median maize yield change is expected by the middle and end of the century under RCP8.5, except in the Congo Basin (central Africa) and in the Ethiopian highlands (eastern Africa), where the median changes in maize yields relatively increase. However, the 95 % confidence interval of the median maize yield changes under both scenarios is broad, and there are opposite signs in most parts of the region. A broad confidence interval implies that the model is subject to many uncertainties. However, in some parts of West Africa, under the RCP8.5 scenario, the confidential interval is finer.

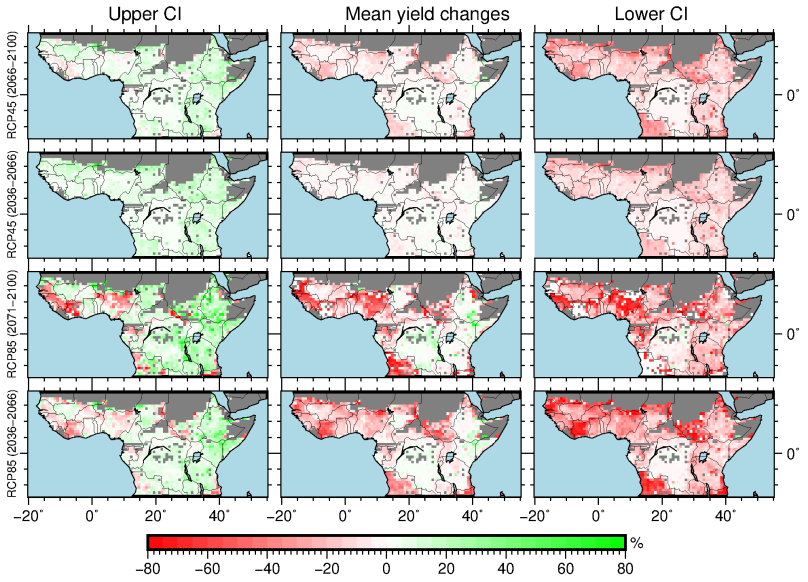

Figure 8.15: Projected median of maize yield changes under RCP4.5 and RCP8.5 during the middle (2036–2066) and end (2071–2100) of the century. The uncertainties of the projection are expressed in the upper and lower 95 % percent confidence intervals.

8.3.2 Multi-Model averaging techniques and the future Maize yield change

The multi-model averaging technique is important in assessing the uncertainties of the statistical model. The statistical model is robust if different multi-model averaging techniques are convergent. Figure 8.16 displays how the six multi-model averaging techniques influence the forecasted maize yield under the RCP8.5 scenario. Firstly, all the techniques indicate a reduction in maize yield during the middle and end of the century. The multi-model averaging based on the assembled mean and R squared projected a relatively low slope and weak inter-annual variability compared to the multi-model averaging based on AICc, AIC, BIC and BMA, which predicted the maize yield variability similarly.

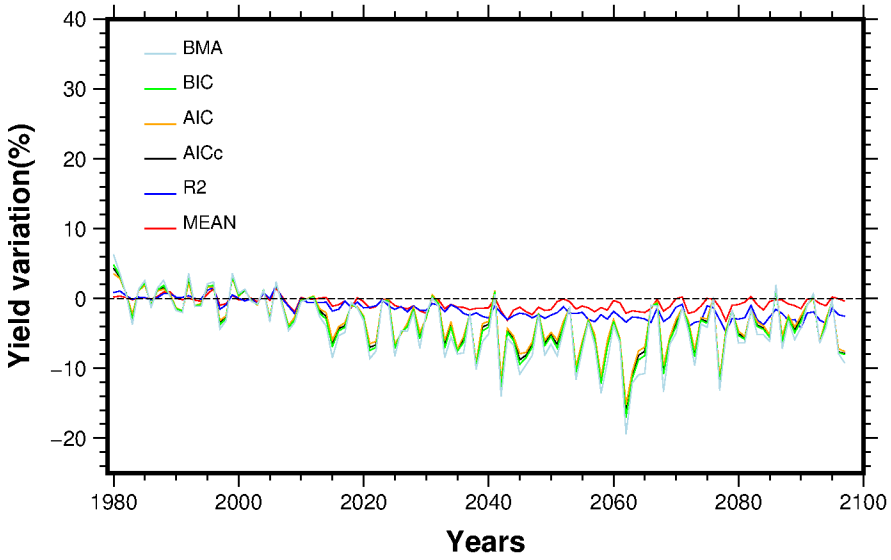

Figure 8.16: Influence of multi-model averaging techniques on the projected maize yield changes.

8.3.3 Separated effects of predictors/groups of predictors and future Maize yield Changes

Different climate predictors have been grouped. These are large-scale predictors, processed-based and traditional local predictors. The influence of these groups of predictors and individual predictors on future maize yield change are presented in Figure 8.17. When

analysed as a group, the maize yield changes are more driven by the large-scale predictors because they almost reproduce the temporal pattern of the yield changes induced by all predictors (Figure 8.17a). The influence of local scale predictors on yield changes is mitigated (Figure 8.17d, 8.17c). However, traditional predictors (Figure 8.17c) seem to be more important in forecasting the future maize yield change than process-based ones (Figure 8.17d).

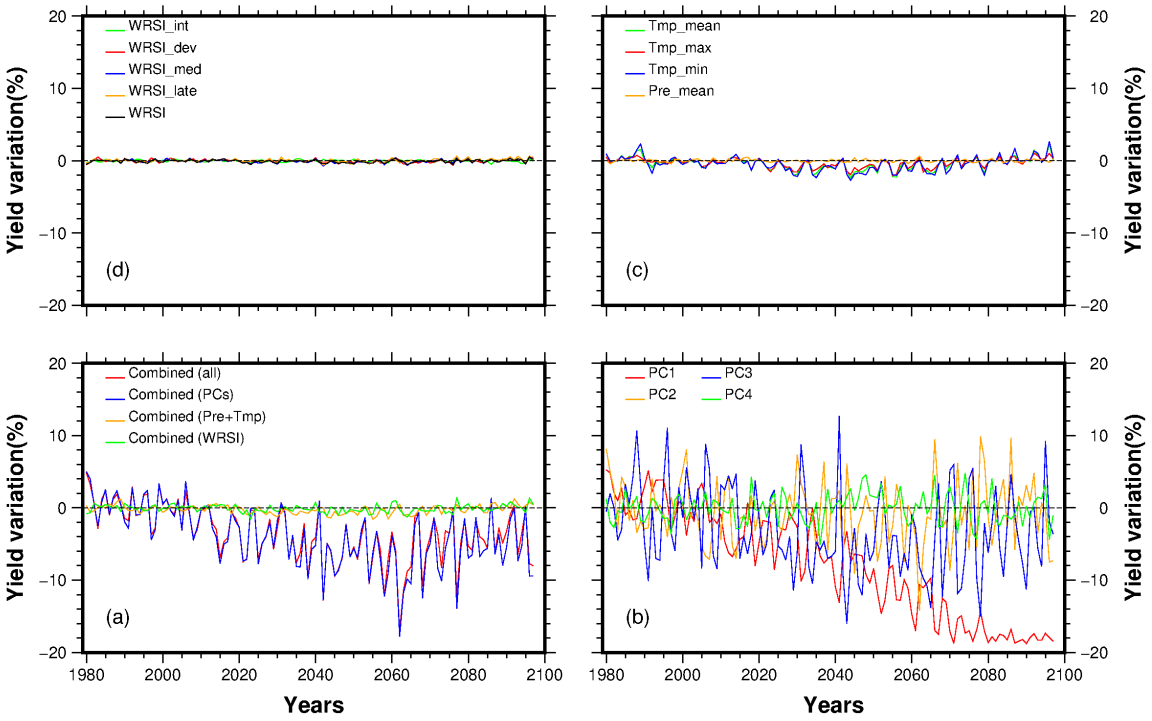

Figure 8.17: Separated effect of a group predictors on future maize yield projection : (a) effects of group of predictors; (b) large scales predictors; (c) traditional local predictors; and (d) local WRSI predictors.

8.4 Projected of WRSI under climate change conditions

As previously stated, the WRSIs used in the statistical crop model were all detrended. The one used here is not detrended and only concerns WRSI-glo. Figure 8.18 shows the spatial distribution of the projected WRSI-glo changes relative to the historical WRSI-glo simulation mean (1982–2005). The results indicate that in the middle of the century, the

WRSI-glo do not substantially change relative to the baseline mean; the magnitude of change is in order of -5 %, reaching 10 % in some parts of the region. The direction of the changes is not clearly depicted; some parts experience a positive change while others experience a negative change. A clear decrease in WRSI-glo is observed at the end of the century; the magnitude of change is around 10 % under RCP4.5 and around 20 % under RCP8.5, globally. However, a significant decrease under RCP8.5 is observed in Sahel and in the Congo Basin. In East Africa,

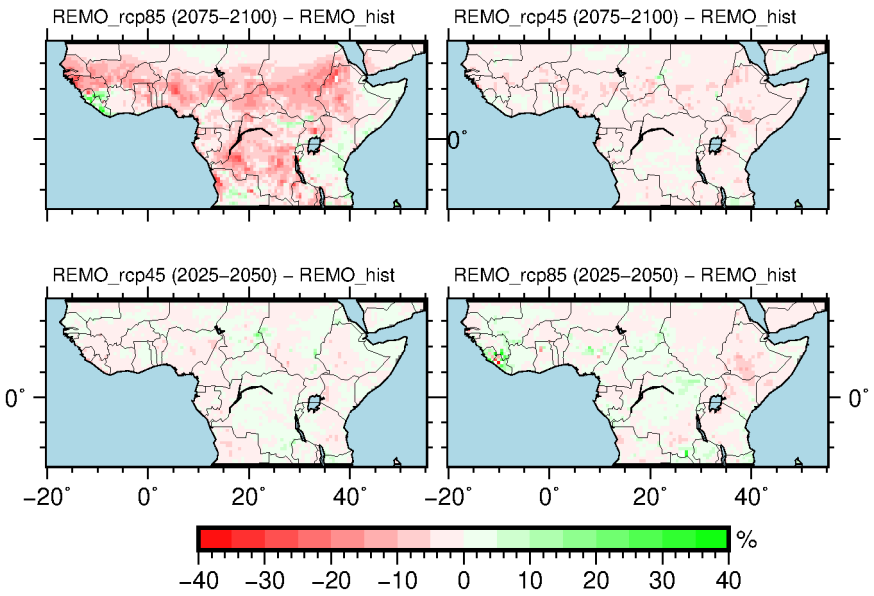

Figure 8.18: Spatial distribution of the difference between the historical WRSI-glo and future projected WRSI-glo under two emission scenarios (RCP4.5 and RCP8.5), by the middle and end of the century.

The simulated times series WRSI-glo under the RCP8.5 and RCP4.5 emission scenarios and the observed ERA5 are presented in Figure 8.19. The visual inspection and the Mann–Kendall test indicate a significant decrease of the available water for maize ($p < 0.001$) over the entirety of tropical Africa under both emission scenarios (Figure 8.19.a). By the end of the century, the overall WRSI-glo will be below the critical value, according to the RCP4.5 emission scenario. The probable change points under RCP4.5 and RCP8.5 are 2067 and 2048, respectively. For a more detailed analysis, a decrease is also observed over central Africa (Figure 8.19b) but is much stronger in RCP8.5 ($p < 0.001$). However, the WRSI-glo is still above the critical value by the end of the century. The changing points under RCP8.5 and RCP4.5 are 2047 and 2020, respectively. In East Africa (Figure

8.19c), a decreasing trend is observed, but is not statistically significant at the 5 % significance level. The simulated times series of the WRSI-glo in West Africa (Figure 8.19d) is quite similar to the one observed in central Africa (Figure 8.19b), characterised by a significant decreasing trend and a relatively high value of WRSI-glo. The changing points of WRSI-glo in West Africa under RCP8.5 and RCP4.5 are 2068 and 2040, respectively.

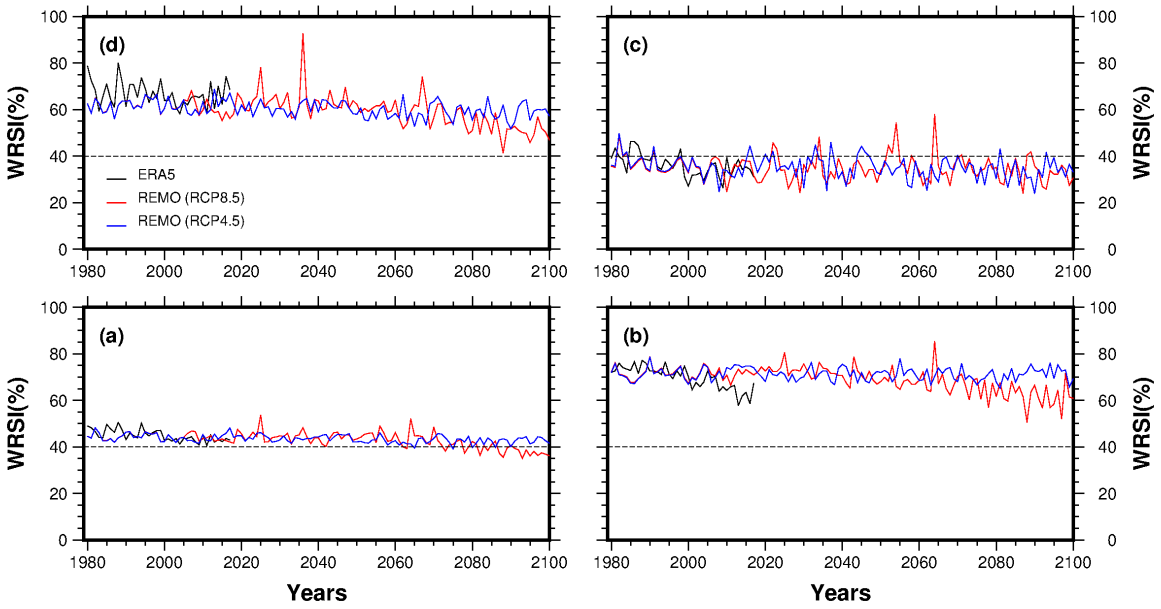

Figure 8.19: Projected WRSI under two climate emission scenarios (RCP8.5 and RCP4.5): (a) tropical Africa – whole study region; (b) central parts of the study region – the Congo Basin; (c) eastern parts of the study region; (d) western parts of the study region.

9. Discussion and recommendations

In this chapter, the key findings of this study are discussed regarding its main objective. These results are compared to other published literature. The limitations of this study and recommendations for further improvement are also addressed. The main objective of this study was to investigate the impact of future climate change and variability on maize yield change using a cross-validated regression. The present-day climate predictors and maize yield data are related using a robust cross-validated multiple linear regression. The results indicate that the regression model performs quite well and confirms that maize yield is largely influenced by the climate variability in tropical Africa. Independent to the adaptation strategies and technologies improvement, climate variability explains, globally, about 27 % of maize yield inter-annual variability in present-day conditions. This finding is similar to the explained variance reported by Ray et al. (2015), although their study was restricted to local predictors such as precipitation and temperature. This study also reveals that climate fluctuation could explain up to 55 % of the year-to-year maize yield variability in the west of Africa. This range of explained variance was also reported by Paeth et al. (2008), who demonstrated that climate variables – mainly temperature, relative humidity and precipitation – could explain about 50 % of the yield variability in Benin. Awoye et al. (2017) also reported the same range of explainted variance when modelling the impact of climate change on maize production using our regression model and large-scale climate predictors (PCs).

The contribution of each predictor or group of predictors in explaining maize yield variability differs. The PCs predictors alone explain approximately 25 % of the total maize yield. The PC2 – which is highly correlated with ENSO indices – was the most important predictor, especially in West Africa, Angola and around the Great Lakes region. The influence of ENSO in West Africa has been much documented (Mohino et al., 2011; Losada et al., 2012), explaining between 24 and 29 % of the total precipitation of the region. Its impact on maize yield variability has also been reported, but at national and local levels in many published papers (Hansen et al., 2004; Iizumi et al., 2014). A strong connection

between maize yield and ENSO has been observed in Zimbabwe and Kenya Hansen et al. (2004). This spatial pattern of the explained variance aligns with the findings of Iizumi and Sakai (2020), who used an older version of our maize yield dataset. They found three sites where ENSO significantly impacted maize yield: West Africa, the Great Lakes region and Angola. This study shows furthermore that the relationship between ENSO and maize yield is globally negative. The same result was observed recently by Iizumi and Sakai (2020). To the best of our knowledge, the impact of other climate teleconnection patterns or climate indices – captured by other PCs – on maize yield has not yet been investigated in a large region. However, few studies havw shown the impact of these indices on either precipitation or temperature variability, implying that they could also impact crop yield variability in tropical Africa. According to Pachauri et al. (2014), SAOD, which has an impact on precipitation, has been assessed in West Africa through numerical model simulations (Kucharski et al., 2009; Losada et al., 2010). Similar, Worou et al. (2020) demonstrated that the TSA – which is significantly correlated with PC2 and PC1 – had an impact on the precipitation in West Africa. IOD, which is also correlated to PC3 and PC4, is highly dominant in East Africa, where it explains up to 80 % of the rainy season precipitation variability and may have an impact on maize yield in these regions (Srivastava et al., 2019b).

As demonstrated, local predictors (traditional and process-based) do not influence interannual maize yield variability much. Still their contribution is more pronounced in the eastern parts, particularly around the Ethiopian highlands. Among local predictors, the mean temperature seems to be the most prominent. The influence of the mean temperature on crop production, in general, has been reported by many scientists (Lobell et al., 2008; Sultan et al., 2014; Waha et al., 2013). Waha et al. (2013) demonstrated that the mean temperature was the most determinant factor of maize yield variability in different locations across the world (France, USA, Brazil and Tanzania). Their study was based on 23 process-based models. Lobell et al. (2011) demonstrated that temperature is more important than precipitation in explaining the year-to-year variability of maize production. Waha et al. (2013) also drew a similar conclusion by conducting their research in Guinea. In most areas where the temperature is predominant (in the eastern parts), the relationship between maize yield and the mean temperature was negative. According to Tao and Zhang (2011), this could be explained by the fact that temperature increase often shortens the growing season length, which in turn reduces the total CO_2 assimilation. Consequently, the total biomass and the grain yield will be reduced. In addition, high temperatures affect the production of pollen by delaying female reproductive tissues,

resulting in low quantity and poor quality of maize grain (Cicchino et al., 2010; Cairns et al., 2013). However, it is important to mention that the relationship between temperature and maize production or other crop yields is not linear (Lobell and Field, 2007). Lobell and Field (2007) indicated that only temperatures above 30 $°C$ could affect maize yield because the optimal temperature is around 29 $°C$. They concluded that each day that experienced a temperature of 30 $°C$ or higher generally resulted in a 1 % loss in maize yield.

To a lesser extent than temperature, the impact of precipitation on inter-annual maize variability is also significant around the Ethiopian highlands. It is an even stronger factor than the temperature in some areas. A combined effect of temperature and precipitation in explaining the maize yield variability in East Africa has also been reported by Ray et al. (2015). Generally, maize yield is very vulnerable to water deficits and surplus (Mansouri-Far et al., 2010; Agbossou et al., 2012), especially during the flowering and grain formation stage (Agbossou et al., 2012). The precipitation usually affects maize yield by reducing the number of grains per year, as well as the quality of the grain. The coefficient of the regression indicates that precipitation negatively influences inter-annual maize yields in eastern Africa. However, in many studies, this relationship is positive. Akponikpe (1999), for example, found that there was maize yield reduction due to reduced precipitation in West Africa. The negative relationship, as observed in this study, between precipitation and maize yield in the highlands of Ethiopia could be explained by the abundance of precipitation in the region, estimated at 3000 $mm/year$. Therefore, floods – which are likely to occur in this region – constitute a major threat to maize production compared to droughts. According to UNOCHA (2017), areas with too much precipitation, such as the highlands of Ethiopia, are expected to see agricultural production be negatively affected by severe flooding. Malawi – where precipitation is the most dominant predictor in this study – has observed 40 flood disasters between 1976 and 2009. The rapid overflowing of Lake Victoria due to extreme precipitation has been recently reported by Wainwright et al. (2021). Several climatic phenomena are responsible for extreme precipitation over East Africa. Black (2005) showed that the El Niño event was responsible for extreme precipitation in East Africa in 1997. Positive IOD is also associated with heavy precipitation in East Africa (Saji et al., 1999). A part from these two major phenomena, the MJO and the tropical cyclone in the western Indian ocean are also related to heavy precipitation in East Africa (Finney et al., 2020).

The results also indicate that WRSI predictors did not influence maize yield variability in the statistical model. This could purely be a statistical issue, as the five WRSIs are

highly correlated with traditional predictors: the cross-validated regression used in this study is likely to select one variable when two or more are highly correlated. Also, the WRSIs predictors may lose much information by being detrended.

We, therefore, assessed directly the impact of climate change on WRSI-glo (WRSI commonly used for crop monitoring and yield estimation). We first correlated it with the observed maize yield; both variables are not detrended. Figure 8.3 and reveals that the correlation is positive in East Africa, reaching 0.6 in some places. Many studies reported a strong positive correlation between WRSI-glo and maize yield in East Africa, precisely in Kenya, Malawi and Mozambique (Jayanthi and Husak, 2013; Jayanthi et al., 2014). A high correlation coefficient was reported between WRSI-glo and maize yield in 174 districts of Ethiopia by Senay and Verdin (2003). More recently, Tarnavsky et al. (2018) found that WRSI-glo was significantly correlated (0.61) with maize yield in Tanzania. The negative correlation coefficient between WRSI-glo and maize yield in West Africa and tropical Africa can be explained by the fact that these regions are less prone to drought, and the maize yield is more affected by water surplus rather than water deficit. As far as the impact of future climate change is concerned, the spatial distribution of the difference between the present-day and the future simulation of WRSI-glo under different scenarios indicates that WRSI-glo is expected to decrease under the RCP8.5 scenario by the end of the century in the entire study area, except in some locations in the east. This result is confirmed by the simulated times series of the WRSI-glo. Crespo et al. (2011) also reported a decrease in WRSI-glo in southern Africa, except in the southeast of the region where a slight increase was observed. According to Burney and Naylor (2012), crop yield failure is caused more by water stress, increased evapotranspiration and precipitation variability than heat stress. Based on this, we can suggest the politics of the construction and the rehabilitation of the irrigation scheme in central Africa. The implementation of large irrigation scales is very challenging and expensive. It is not evident that local farmers will adopt irrigation as coping strategy because of the affordability. In addition, open water is becoming less available due to high demand for sector competing sectors: hydroelectricity and urbanisation. In this regard, our suggestion would be to focus on small irrigation scales.

As far as the future maize yield change is concerned, this study suggested an overall decrease of maize yield by the end of the century – around 5% under RCP4.5 and 20 % under RCP8.5 in the middle of the century. The analysis of the spatial distribution indicates that the median maize yield has increased everywhere, except in the Congo basin and East Africa, particularly in the Ethiopian highlands. Maize yield reduction across

many tropical African countries has been projected by many other scientists (Jones and Thornton, 2003). The majority of studies indicate a general decrease in maize yield over SSA (Abate et al., 2015). Thornton et al. (2009) use GCM climate projections to assess the impact of climate change on maize yield; they reported that the projected maize yield change was spatially heterogeneous: large areas were experiencing reduced yield change and gain maize yields were projected in East Africa. Thornton et al. (2011) also projected a moderate increase of maize yield in East Africa, about 18 % in Kenya and 18 % in Rwanda, by 2050. However, these results can vary according to the climate model and maize yield datasets. Knox et al. (2012) assessed 50 studies and concluded that the impact of climate change on maize yield might vary from -50 to +100 %. This study also demonstrates that climate emission scenarios have an impact on future maize yield change. We observe a significant decrease in median maize yield change from 1982 to 2065, and an increase after 2066 up to the end of the century. However, this trend is not observed in the RCP4.5 emission scenario. The impact of the climate emission scenario is also well depicted on the spatial pattern of the maize yield change, where a remarkable decrease in maize yield is observed everywhere under RCP8.5, except in some areas of eastern and central Africa. These findings are very intuitive because RCPs emission scenarios are developed with different proportions of land cover type and concentration of greenhouse gases. These two factors are major causes of climate change in tropical Africa (Douville et al., 2000; Feddema and Freire, 2001; Lu and Delworth, 2005; Knorr and Schnitzler, 2006; Paeth et al., 2009); thus, they are related to crop production. Under RCP8.5, croplands and grasslands are the most dominant land-use types, while natural vegetation is predominant under RCP4.5, as indicated in Figure 9.1

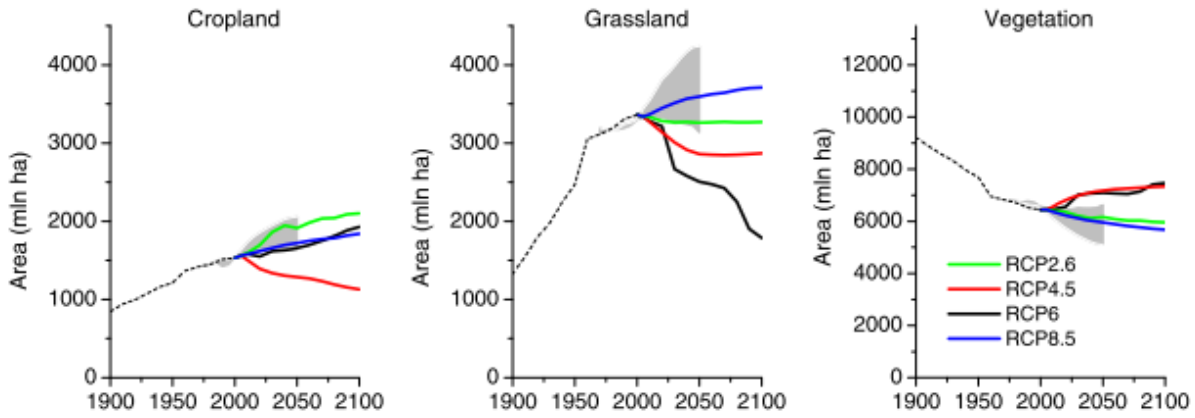

Figure 9.1: Vegetation crop land use and grass land use across RCPs. The gray areas denote the 95 percentiles scenario (source: Van Vuuren et al., 2011).

According to Gaiser et al. (2011), land cover degradation had already led to 24 % maize yield reduction in Benin. The same authors also concluded that the situation was likely to be amplified in the coming decades if no serious measures are taken. This result was later confirmed by Challinor et al. (2015). Land-use changes are indirectly related to crop production by increasing the temperature and disturbing evapotranspiration. According to Paeth et al. (2009), land-use change is responsible for about 35 % of the mean temperature over sub-Saharan Africa. Moore et al. (2014) also showed that land-use changes also had an impact on local precipitation by modifying the cloud physics. In West Africa, for example, deforestation is considered as the main cause of drought (Clark et al., 2001; Abiodun et al., 2008; Paeth et al., 2008, 2009). The degradation of tropical forests alone is responsible for approximately 26 % of the global CO_2 in the atmosphere (Ramankutty et al., 2008). A difference between RCP8.5 and RCP4.5 is also attributed to the concentration of CO_2. RCP8.5 is driven by the CO_2 equivalent of almost 1370 *ppm*, while RCP4.5 is driven by the CO_2 equivalent of about 850 *ppm*. However, as a C4 plant, maize is less sensitive to an increased CO_2 because photosynthesis is saturated at current CO_2 concentrations (Leakey et al., 2009; Cairns et al., 2013; Parkes et al., 2018). In contrast to C4 plants, the increase of CO_2 in the atmosphere is likely to simulate photosynthesis for C3 crops (*i.e.* rice, wheat) by improving the carboxylation rate of rubisco, thus increasing production (Ainsworth and Long, 2005). We can conclude that the difference between RCP8.5 and RCP4.5 could be much attributed to land use-changes, as argued above.

This study does not address all the complexities related to climate and crop yield modelling. Its results are stickily related to the proposed statistical model, type of protectors and datasets. However, it can serve as a solid basis for further investigations. One of the most prominent weaknesses of this study is the non-integration of adaptation strategies. The study assumes that farmers have no possibility of adapting to the impact of climate change. Moreover, the technology improvements are not counted. The result could largely differ if climate adaptation options and technology improvements were part of the model because African farmers are likely to mitigate the negative influence of climate variability on agriculture production by adopting many agricultural coping strategies, such as the mixed crops (Butt et al., 2005; Roudier et al., 2011; Rurinda et al., 2014). Unfortunately, adaptation options could only be integrated in a process-based modelling approach (Rosenzweig et al., 2014; Caubel et al., 2015). Apart from the adaptation strategies, the statistical model, as developed in this study, could not capture the possible non-linear relationship between of climate predictors and the maize yield. According to Schlenker and

Roberts (2009) and Lobell et al. (2011), the relationship between temperature and crop production is often non-linear. Generally, non- linear relationships are modelled in regression by adding a quadratic term (Schlenker and Lobell, 2010; Blanc, 2012). However, this is possible with a regression with fixed predictors. The cross-validated regression, as implemented in this study, would not select both the variable and its corresponding quadratic term because they are highly correlated. Novel predictive models based on machine learning and deep learning approaches could address this problem. These methods can to deal with complex datasets and generally outperform parametric regressions in predicting out-of-sample datasets. Crane-Droesch (2018) tested the performance of semiparametric variants of deep neural network – which is said to be the best predictive model – in modelling crop yields against a multiple linear regression. The results indicated that the semiparametric neural network could simultaneously model the complex non-linear relationship and the parametric structure of the dataset and was far better than multiple linear regression. This study also omitted integrating factors that could influence crop yields, such as weeds, pests and parasite plants. Many researchers have projected an increase in these factors with the current climate change trend, and future crop yields are likely to be negatively affected (Fand et al., 2012; Knox et al., 2012; Fand et al., 2014). Based on these limitations, particularly technological improvement and the non-linearity, it will be interesting if the future development of the crop yield model is based on semiparametric variants of deep neural network. It will also be interesting if the technological improvement is added to the empirical model using the method suggested by (Zhai et al., 2021). They defined technological improvement as linear fitting and transferred it to the future maize yield change.

The uncertainties linked to datasets constitute a threat to this study. Both the crop yield dataset and datasets used to generate predictors are subjects to huge uncertainties. According to many scientists, climate models do not perform well in tropical Africa, especially in equatorial Africa, because of the poor quality of observational datasets (Baccini et al., 2012; Panitz et al., 2016). Climate model uncertainties can derive from many sources: initial conditions, boundary condition uncertainties, parameters and structural uncertainties (Tebaldi and Knutti, 2007). Precipitation, for example, has been projected to increase according to some models but not to others (Haensler et al., 2013; Lawrence and Vandecar, 2015). Concerning the temperature, almost all climate models agree with the projected direction of change, but the magnitude is still controversial. Consequently, the uncertainties of climate models are transferred to future maize yield changes. However, it is necessary to point out that the regional climate model REMO used in this

study performed quite well and can reproduce important variables such as the precipitation and temperature (Fotso-Nguemo et al., 2017a). Nevertheless, the multiple climate models that use many emission scenarios could have been the most efficient methods to reduce the uncertainties of climate prediction (Lobell et al., 2006; Mishra and Singh, 2010). If climate model uncertainties were easily addressed by assembling models (Tao et al., 2008; Hawkins and Sutton, 2011), scientists have paid less attention in improving uncertainties linked to crop data (Lobell et al., 2006). Few studies addressed this issue through the Agricultural Model Intercomparison and Improvement Project (AgMIP) framework. Still, it is only possible with process-based crop models. Little is said about the uncertainties linked to historically observed crop data. However, Anderson et al. (2015) compared four cropping systems and concluded that they were significantly different. More recently, Iizumi et al. (2017) assessed the uncertainties of the maize yield used in this study by comparing it with the FAO census historical crop yields. Their results indicated that both datasets matched well in yield trend, and the spatial variability of GDHY datasets – also used in this study – was much better. The uncertainties of GDHY mainly result from (1) the yield gaps and uncertainties of the FAO census dataset, which is the primary product; (2) the time-constant information on harvested area; and (3) the uncertainties of the satellite dataset. To address some sources of uncertainty, we implemented a spatial smoothing of the nine-neighbouring grid because the data comes from different spatial resolutions. In addition, the multi-model averaging was performed, and four of them yielded similar results. However, in the end, we used the BMA approach because it was the most performant. Awoye et al. (2017) also reported the high performance of all Bayesian based multi-model averaging techniques.

With respect to what is demonstrated in this study, some practical recommendations must be formulated for decision-makers. It is clearly demonstrated that the production of maize – which is the most important crop in Africa – will be negatively affected by climate change. The most realistic decision could be to expand the agriculture areas, but this option is not efficient, regarding the expansion of urbanisation, the need for afforestation and other non-agricultural activities. We will therefore recommend the three strategies suggested by FAO (2016): (1) the reduction of post-harvest losses by investing in agricultural roads, logistics, storage and processing; (2) agriculture intensification; and (3) the promotion of multiple cropping systems, which will increase the number of cropping seasons. Further adaptation strategies, depending on the context of each region, are also welcome. This could be in the form of income management, government assurance and programes, farms production practices and technology development. As

farmers are particularly affected by climate change, more action should be directed towards them. The advantages and limitations of farm production adaptation options have been addressed by many scientists (Sultan and Gaetani, 2016; Guan et al., 2017; Parkes et al., 2018). Amongst a large variety of realistic adaptation strategies, the planting of maize cultivars with thermal time requirements and greater resilience to heat stress is the most promising approach and is highly recommended by this study. Crop density, late sowing and mixing crop could also be applied, according to the context of the study area. However, these options should be implemented after considering existing adaptation strategies implemented by the local community, as they have already started developing their own coping strategy (Kurukulasuriya et al., 2006). According to (Kristjanson et al., 2012), marginal adaptation strategies, such as crop variety and planting date, have already been successfully implemented by many farmers in tropical Africa. However, these methods are still limited in improving the soil, acquiring adequate water, and adopting land management practices.

Bibliography

Abate, T., Shiferaw, B., Menkir, A., Wegary, D., Kebede, Y., Tesfaye, K., Kassie, M., Bogale, G., Tadesse, B., and Keno, T. (2015). Factors that transformed maize productivity in ethiopia. *Food security*, 7(5):965–981.

Abiodun, B., Pal, J., Afiesimama, E., Gutowski, W., and Adedoyin, A. (2008). Simulation of west african monsoon using regcm3 part ii: impacts of deforestation and desertification. *Theoretical and Applied Climatology*, 93(3):245–261.

Abiye, O. E., Matthew, O. J., Sunmonu, L. A., and Babatunde, O. A. (2019). Potential evapotranspiration trends in west africa from 1906 to 2015. *SN Applied Sciences*, 1(11):1–14.

Adejuwon, J. O. (2006). Food crop production in nigeria. ii. potential effects of climate change. *Climate Research*, 32(3):229–245.

Adhikari, P., Araya, H., Aruna, G., Balamatti, A., Banerjee, S., Baskaran, P., Barah, B., Behera, D., Berhe, T., Boruah, P., et al. (2018). System of crop intensification for more productive, resource-conserving, climate-resilient, and sustainable agriculture: Experience with diverse crops in varying agroecologies. *International journal of agricultural sustainability*, 16(1):1–28.

Agbossou, E., Toukon, C., Akponikpè, P., and Afouda, A. (2012). Climate variability and implications for maize production in benin: A stochastic rainfall analysis. *African Crop Science Journal*, 20:493–503.

Ahmed, K. F., Wang, G., Silander, J., Wilson, A. M., Allen, J. M., Horton, R., and Anyah, R. (2013). Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the us northeast. *Global and Planetary Change*, 100:320–332.

Ainsworth, E. A. and Long, S. P. (2005). What have we learned from 15 years of free-air co2 enrichment (face)? a meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising co2. *New phytologist*, 165(2):351–372.

Akinsanola, A., Ogunjobi, K., Ajayi, V., Adefisan, E., Omotosho, J., and Sanogo, S. (2017). Comparison of five gridded precipitation products at climatological scales over west africa. *Meteorology and Atmospheric Physics*, 129(6):669–689.

Aleman, J. C., Jarzyna, M. A., and Staver, A. C. (2018). Forest extent and deforestation in tropical africa since 1900. *Nature ecology & evolution*, 2(1):26–33.

Alexandrov, V. and Hoogenboom, G. (2000). The impact of climate variability and change on crop yield in bulgaria. *Agricultural and forest meteorology*, 104(4):315–327.

Ali, A., Lebel, T., and Amani, A. (2005). Rainfall estimation in the sahel. part i: Error function. *Journal of Applied Meteorology*, 44(11):1691–1706.

Allen, R. G., Pereira, L. S., Howell, T. A., and Jensen, M. E. (2011). Evapotranspiration information reporting: Ii. recommended documentation. *Agricultural Water Management*, 98(6):921–929.

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al. (1998). Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *Fao, Rome*, 300(9):D05109.

Almazroui, M., Saeed, S., Saeed, F., Islam, M. N., and Ismail, M. (2020). Projections of precipitation and temperature over the south asian countries in cmip6. *Earth Systems and Environment*, 4(2):297–320.

Aloysius, N. R., Sheffield, J., Saiers, J. E., Li, H., and Wood, E. F. (2016). Evaluation of historical and future simulations of precipitation and temperature in central africa from cmip5 climate models. *Journal of Geophysical Research: Atmospheres*, 121(1):130–152.

Amadu, F. O., Miller, D. C., and McNamara, P. E. (2020). Agroforestry as a pathway to agricultural yield impacts in climate-smart agriculture investments: Evidence from southern malawi. *Ecological Economics*, 167:106443.

Amondo, E. and Simtowe, F. (2018). Technology innovations, productivity and production risk effects of adopting drought tolerant maize varieties in rural zambia. Technical report.

Anderson, W., You, L., Wood, S., Wood-Sichra, U., and Wu, W. (2015). An analysis of methodological and spatial differences in global cropping systems models and maps. *Global Ecology and Biogeography*, 24(2):180–191.

Anyanwu, J. C., Anyanwu, J. C., et al. (2017). The key drivers of poverty in sub-saharan africa and what can be done about it to achieve the poverty sustainable development goal. *Asian Journal of Economic Modelling*, 5(3):297–317.

Ashaolu, E. D. and Iroye, K. (2018). Rainfall and potential evapotranspiration patterns and their effects on climatic water balance in the western lithoral hydrological zone of nigeria. *Ruhuna Journal of Science*, 9(2).

Awoye, O., Pollinger, F., Agbossou, E., and Paeth, H. (2017). Dynamical-statistical projections of the climate change impact on agricultural production in benin by means of a cross-validated linear model combined with bayesian statistics. *Agricultural and Forest Meteorology*, 234:80–94.

Baccini, A., Goetz, S., Walker, W., Laporte, N., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P., Dubayah, R., Friedl, M., et al. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature climate change*, 2(3):182–185.

Bahaga, T. K., Fink, A. H., and Knippertz, P. (2019). Revisiting interannual to decadal teleconnections influencing seasonal rainfall in the greater horn of africa during the 20th century. *International Journal of Climatology*, 39(5):2765–2785.

BAHRENBERG, G., GIESE, E., and NIPPER, J. (1999). Statistische methoden in der geographie; 4. überarbeitete auflage. *Stuttgart. Leipzig.*

Bala, G., Caldeira, K., Wickett, M., Phillips, T., Lobell, D., Delire, C., and Mirin, A. (2007). Combined climate and carbon-cycle effects of large-scale deforestation. *Proceedings of the National Academy of Sciences*, 104(16):6550–6555.

Balek, J. (2011). *Hydrology and water resources in tropical Africa*. Elsevier.

Barnston, A. G. and Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly weather review*, 115(6):1083–1126.

Barros, V. R., Field, C. B., et al. (2014). Climate change 2014: impacts, adaptation, and vulnerability. part b: regional aspects.

Benestad, R. (2004). Empirical-statistical downscaling in climate modeling. *Eos, Transactions American Geophysical Union*, 85(42):417–422.

Benestad, R. E., Parding, K. M., Isaksen, K., and Mezghani, A. (2016). Climate change and projections for the barents region: what is expected to change and what will stay the same? *Environmental Research Letters*, 11(5):054017.

Bernstein, L., Bosch, P., Canziani, O., Chen, Z., Christ, R., and Riahi, K. (2008). Ipcc, 2007: climate change 2007: synthesis report.

Black, E. (2005). The relationship between indian ocean sea–surface temperature and east african rainfall. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 363(1826):43–47.

Blanc, E. (2012). The impact of climate change on crop yields in sub-saharan africa.

Blanc, É. (2017). Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models. *Agricultural and Forest Meteorology*, 236:145–161.

Blanc, E. and Strobl, E. (2013). The impact of climate change on cropland productivity: evidence from satellite based products at the river basin scale in africa. *Climatic change*, 117(4):873–890.

Blanc, E. and Sultan, B. (2015). Emulating maize yields from global gridded crop models using statistical estimates. *Agricultural and Forest Meteorology*, 214:134–147.

Bombelli, A., Henry, M., Castaldi, S., Adu-Bredu, S., Arneth, A., Grandcourt, A. d., Grieco, E., Kutsch, W. L., Lehsten, V., Rasile, A., et al. (2009). An outlook on the sub-saharan africa carbon balance. *Biogeosciences*, 6(10):2193–2205.

Bortz, J. and Schuster, C. (2011). *Statistik für Human-und Sozialwissenschaftler: Limitierte Sonderausgabe*. Springer-Verlag.

Brink, A. B., Bodart, C., Brodsky, L., Defourney, P., Ernst, C., Donney, F., Lupi, A., and Tuckova, K. (2014). Anthropogenic pressure in east africa—monitoring 20 years of land cover changes by means of medium resolution satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 28:60–69.

Brovkin, V., Raddatz, T., Reick, C. H., Claussen, M., and Gayler, V. (2009). Global biogeophysical interactions between forest and climate. *Geophysical research letters*, 36(7).

Burney, J. A. and Naylor, R. L. (2012). Smallholder irrigation as a poverty alleviation tool in sub-saharan africa. *World Development*, 40(1):110–123.

Burnham, K. P. and Anderson, D. R. (2014). P values are only an index to evidence: 20th-vs. 21st-century statistical science. *Ecology*, 95(3):627–630.

Butt, T. A., McCarl, B. A., Angerer, J., Dyke, P. T., and Stuth, J. W. (2005). The economic and food security implications of climate change in mali. *Climatic change*, 68(3):355–378.

Cairns, J. E., Hellin, J., Sonder, K., Araus, J. L., MacRobert, J. F., Thierfelder, C., and Prasanna, B. M. (2013). Adapting maize production to climate change in sub-saharan africa. *Food Security*, 5(3):345–360.

Camberlin, P. (2018). Climate of eastern africa. In *Oxford Research Encyclopedia of Climate Science*.

Campos, I., Balbontín, C., González-Piqueras, J., Gonzalez-Dugo, M. P., Neale, C. M., and Calera, A. (2016). Combining a water balance model with evapotranspiration measurements to estimate total available soil water in irrigated and rainfed vineyards. *Agricultural Water Management*, 165:141–152.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q. (2015). Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, 28(17):6938–6959.

Carter, T., Alfsen, K., Barrow, E., Bass, B., Dai, X., Desanker, P., Gaffin, S., Giorgi, F., Hulme, M., Lal, M., et al. (2007). General guidelines on the use of scenario data for climate impact and adaptation assessment. *Finnish Environmental Institute, Helsinki, Finland*.

Caubel, J., de Cortázar-Atauri, I. G., Launay, M., de Noblet-Ducoudré, N., Huard, F., Bertuzzi, P., and Graux, A.-I. (2015). Broadening the scope for ecoclimatic indicators to assess crop climate suitability according to ecophysiological, technical and quality criteria. *Agricultural and Forest Meteorology*, 207:94–106.

Challinor, A. J., Parkes, B., and Ramirez-Villegas, J. (2015). Crop yield response to climate change varies with cropping intensity. *Global change biology*, 21(4):1679–1688.

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S., Smith, D., and Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, 4(4):287–291.

Change, C. (1992). The supplementary report to the ipcc scientific assessment. *Intergovernmental Panel on Cllmate Change, Cambrldge Univ Press, Cambridge*.

Chauvin, N. D., Mulangu, F., Porto, G., et al. (2012). Food production and consumption trends in sub-saharan africa: Prospects for the transformation of the agricultural sector. *UNDP Regional Bureau for Africa: New York, NY, USA*.

Chen, C., Haerter, J. O., Hagemann, S., and Piani, C. (2011). On the contribution of statistical bias correction to the uncertainty in the projected hydrological cycle. *Geophysical Research Letters*, 38(20).

Christensen, J. H., Boberg, F., Christensen, O. B., and Lucas-Picher, P. (2008). On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters*, 35(20).

Christy, J. R., Norris, W. B., and McNider, R. T. (2009). Surface temperature variations in east africa and possible causes. *Journal of Climate*, 22(12):3342–3356.

Cicchino, M., Edreira, J. R., Uribelarrea, M., and Otegui, M. (2010). Heat stress in field-grown maize: Response of physiological determinants of grain yield. *Crop science*, 50(4):1438–1448.

Cionni, I., Eyring, V., Lamarque, J.-F., Randel, W., Stevenson, D., Wu, F., Bodeker, G., Shepherd, T., Shindell, D., and Waugh, D. (2011). Ozone database in support of cmip5 simulations: results and corresponding radiative forcing. *Atmospheric Chemistry and Physics*, 11(21):11267–11292.

Clark, D. A., Brown, S., Kicklighter, D. W., Chambers, J. Q., Thomlinson, J. R., Ni, J., and Holland, E. A. (2001). Net primary production in tropical forests: an evaluation and synthesis of existing field data. *Ecological applications*, 11(2):371–384.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., et al. (2013). Long-term climate change:

projections, commitments and irreversibility. In *Climate Change 2013-The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1029–1136. Cambridge University Press.

Conway, D., Van Garderen, E. A., Deryng, D., Dorling, S., Krueger, T., Landman, W., Lankford, B., Lebek, K., Osborn, T., Ringler, C., et al. (2015). Climate and southern africa's water–energy–food nexus. *Nature Climate Change*, 5(9):837–846.

Cook, K. H. (1999). Generation of the african easterly jet and its role in determining west african precipitation. *Journal of climate*, 12(5):1165–1184.

Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11):114003.

Crespo, O., Hachigonta, S., and Tadross, M. (2011). Sensitivity of southern african maize yields to the definition of sowing dekad in a changing climate. *Climatic Change*, 106(2):267–283.

Davenport, F., Funk, C., and Galu, G. (2018). How will east african maize yields respond to climate change and can agricultural development mitigate this response? *Climatic change*, 147(3):491–506.

Deryng, D., Conway, D., Ramankutty, N., Price, J., and Warren, R. (2014). Global crop yield response to extreme heat stress under multiple climate change futures. *Environmental Research Letters*, 9(3):034011.

Di Luca, A., de Elía, R., and Laprise, R. (2015). Challenges in the quest for added value of regional climate dynamical downscaling. *Current Climate Change Reports*, 1(1):10–21.

Diem, J. E., Ryan, S. J., Hartter, J., and Palace, M. W. (2014). Satellite-based rainfall data reveal a recent drying trend in central equatorial africa. *Climatic Change*, 126(1):263–272.

Dinku, T., Hailemariam, K., Maidment, R., Tarnavsky, E., and Connor, S. (2014). Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over ethiopia. *International Journal of Climatology*, 34(7):2489–2504.

Dlamini, L. (2013). *Modelling of standardised precipitation index using remote sensing for improved drought monitoring.* PhD thesis.

Dobler, A. and Ahrens, B. (2008). Precipitation by a regional climate model and bias correction in europe and south asia. *Meteorologische Zeitschrift*, pages 499–509.

Doorenbos, J. and Pruitt, W. (1977). Crop water requirements, fao irrigation and drainage. *Paper 24, 2a ed., Roma, Italy.*

Dosio, A. and Panitz, H.-J. (2016). Climate change projections for cordex-africa with cosmo-clm regional climate model and differences with the driving global climate models. *Climate Dynamics*, 46(5):1599–1625.

Dosio, A., Panitz, H.-J., Schubert-Frisius, M., and Lüthi, D. (2015). Dynamical downscaling of cmip5 global circulation models over cordex-africa with cosmo-clm: evaluation over the present climate and analysis of the added value. *Climate Dynamics*, 44(9):2637–2661.

Douville, H., Planton, S., Royer, J.-F., Stephenson, D., Tyteca, S., Kergoat, L., Lafont, S., and Betts, R. (2000). Importance of vegetation feedbacks in doubled-co2 climate experiments. *Journal of Geophysical Research: Atmospheres*, 105(D11):14841–14861.

Dube, S., Scholes, R. J., Nelson, G. C., Mason-D'Croz, D., and Palazzo, A. (2013). South african food security and climate change: Agriculture futures. *Economics: The Open-Access, Open-Assessment E-Journal*, 7(2013-35):1–54.

Dunning, C. M., Black, E., and Allan, R. P. (2018). Later wet seasons with more intense rainfall over africa under future climate change. *Journal of Climate*, 31(23):9719–9738.

Enayati, M., Bozorg-Haddad, O., Bazrafshan, J., Hejabi, S., and Chu, X. (2021). Bias correction capabilities of quantile mapping methods for rainfall and temperature variables. *Journal of Water and Climate Change*, 12(2):401–419.

Engelbrecht, F., Adegoke, J., Bopape, M.-J., Naidoo, M., Garland, R., Thatcher, M., McGregor, J., Katzfey, J., Werner, M., Ichoku, C., et al. (2015). Projections of rapidly rising surface temperatures over africa under low mitigation. *Environmental Research Letters*, 10(8):085004.

Falconnier, G. N., Corbeels, M., Boote, K. J., Affholder, F., Adam, M., MacCarthy, D. S., Ruane, A. C., Nendel, C., Whitbread, A. M., Justes, É., et al. (2020). Modelling

climate change impacts on maize yields under low nitrogen input conditions in sub-saharan africa. *Global change biology*, 26(10):5942–5964.

Fand, B. B., Kamble, A. L., and Kumar, M. (2012). Will climate change pose serious threat to crop pest management: A critical review. *International journal of scientific and Research publications*, 2(11):1–14.

Fand, B. B., Tonnang, H. E., Kumar, M., Kamble, A. L., and Bal, S. K. (2014). A temperature-based phenology model for predicting development, survival and population growth potential of the mealybug, phenacoccus solenopsis tinsley (hemiptera: Pseudococcidae). *Crop Protection*, 55:98–108.

FAO, O. (2016). Price volatility in food and agricultural markets: Policy responses. policy report including contributions by fao, ifad, imf.

FAO-ITTO (2011). The state of forests in the amazon basin, congo basin and southeast asia.

Feddema, J. J. and Freire, S. (2001). Soil degradation, global warming and climate impacts. *Climate Research*, 17(2):209–216.

Field, C. B. and Barros, V. R. (2014). *Climate change 2014–Impacts, adaptation and vulnerability: Regional aspects*. Cambridge University Press.

Finney, D. L., Marsham, J. H., Walker, D. P., Birch, C. E., Woodhams, B. J., Jackson, L. S., and Hardy, S. (2020). The effect of westerlies on east african rainfall and the associated role of tropical cyclones and the madden–julian oscillation. *Quarterly Journal of the Royal Meteorological Society*, 146(727):647–664.

Fogt, R. L. and Marshall, G. J. (2020). The southern annular mode: variability, trends, and climate impacts across the southern hemisphere. *Wiley Interdisciplinary Reviews: Climate Change*, 11(4):e652.

Fotso-Nguemo, T. C., Vondou, D. A., Pokam, W. M., Djomou, Z. Y., Diallo, I., Haensler, A., Tchotchou, L. A. D., Kamsu-Tamo, P. H., Gaye, A. T., and Tchawoua, C. (2017a). On the added value of the regional climate model remo in the assessment of climate change signal over central africa. *Climate Dynamics*, 49(11):3813–3838.

Fotso-Nguemo, T. C., Vondou, D. A., Tchawoua, C., and Haensler, A. (2017b). Assessment of simulated rainfall and temperature from the regional climate model remo and future changes over central africa. *Climate Dynamics*, 48(11):3685–3705.

Fouquart, Y. et al. (1980). Computations of solar heating of the earth's atmosphere: A new parameterization.

Frere, M. and Popov, G. (1979). *Agrometeorological crop monitoring and forecasting*. FAO.

Frère, M. and Popov, G. (1986). Early agrometeorological crop yield forecasting. *FAO, Plant Production and Protection Paper*, (73).

Gadgil, S., Vinayachandran, P., Francis, P., and Gadgil, S. (2004). Extremes of the indian summer monsoon rainfall, enso and equatorial indian ocean oscillation. *Geophysical Research Letters*, 31(12).

Gaiser, T., Judex, M., Igué, A. M., Paeth, H., and Hiepe, C. (2011). Future productivity of fallow systems in sub-saharan africa: Is the effect of demographic pressure and fallow reduction more significant than climate change? *Agricultural and Forest Meteorology*, 151(8):1120–1130.

Galmarini, S., Cannon, A. J., Ceglar, A., Christensen, O. B., de Noblet-Ducoudré, N., Dentener, F., Doblas-Reyes, F. J., Dosio, A., Gutierrez, J. M., Iturbide, M., et al. (2019). Adjusting climate model bias for agricultural impact assessment: How to cut the mustard. *Climate services*, 13:65–69.

Gbobaniyi, E., Sarr, A., Sylla, M. B., Diallo, I., Lennard, C., Dosio, A., Dhiédiou, A., Kamga, A., Klutse, N. A. B., Hewitson, B., et al. (2014). Climatology, annual cycle and interannual variability of precipitation and temperature in cordex simulations over west africa. *International Journal of Climatology*, 34(7):2241–2257.

Gebrechorkos, S. H., Hülsmann, S., and Bernhofer, C. (2018). Evaluation of multiple climate data sources for managing environmental resources in east africa. *Hydrology and Earth System Sciences*, 22(8):4547–4564.

Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., Vuuren, D. P. v., Berg, M. v. d., Feng, L., Klein, D., et al. (2019). Global emissions pathways under different socioeconomic scenarios for use in cmip6: a dataset of harmonized emissions trajectories through the end of the century. *Geoscientific model development*, 12(4):1443–1475.

Giorgi, F. and Gutowski Jr, W. J. (2015). Regional dynamical downscaling and the cordex initiative. *Annual Review of Environment and Resources*, 40:467–490.

Giorgi, F., Hewitson, B., Arritt, R., Gutowski, W., Knutson, T., Landsea, C., et al. (2001). Regional climate information—evaluation and projections.

Gleixner, S., Demissie, T., and Diro, G. T. (2020). Did era5 improve temperature and precipitation reanalysis over east africa? *Atmosphere*, 11(9):996.

Guan, K., Sultan, B., Biasutti, M., Baron, C., and Lobell, D. B. (2017). Assessing climate adaptation options and uncertainties for cereal systems in west africa. *Agricultural and Forest Meteorology*, 232:291–305.

Haensler, A., Saeed, F., and Jacob, D. (2013). Assessing the robustness of projected precipitation changes over central africa on the basis of a multitude of global and regional climate projections. *Climatic Change*, 121(2):349–363.

Hall, A. (2014). Projecting regional change. *Science*, 346(6216):1461–1462.

Hansen, J. W., Potgieter, A., and Tippett, M. K. (2004). Using a general circulation model to forecast regional wheat yields in northeast australia. *Agricultural and Forest Meteorology*, 127(1-2):77–92.

Harris, I., Osborn, T. J., Jones, P., and Lister, D. (2020). Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset. *Scientific data*, 7(1):1–18.

Hart, N. C., Washington, R., and Stratton, R. A. (2018). Stronger local overturning in convective-permitting regional climate model improves simulation of the subtropical annual cycle. *Geophysical Research Letters*, 45(20):11–334.

Hartkamp, A. D., De Beurs, K., Stein, A., and White, J. W. (1999). Interpolation techniques for climate variables.

Hartmann, D. L., Tank, A. M. K., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y. A. R., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., et al. (2013). Observations: atmosphere and surface. In *Climate change 2013 the physical science basis: Working group I contribution to the fifth assessment report of the intergovernmental panel on climate change*, pages 159–254. Cambridge University Press.

Hassan, R. M. and Nhemachena, C. (2008). Determinants of african farmers' strategies for adapting to climate change: Multinomial choice analysis. *African Journal of Agricultural and Resource Economics*, 2(311-2016-5521):83–104.

Hastenrath, S. and Greischar, L. (1993). Circulation mechanisms related to northeast brazil rainfall anomalies. *Journal of Geophysical Research: Atmospheres*, 98(D3):5093–5102.

Hawkins, E. and Sutton, R. (2011). The potential to narrow uncertainty in projections of regional precipitation change. *Climate dynamics*, 37(1):407–418.

Hoeffler, H. (2011). The political economy of agricultural policies in africa: History, analytical concepts and implications for development cooperation. *Quarterly Journal of International Agriculture*, 50(892-2016-65192):29–53.

Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., Heng, Y., Konopka, P., Müller, R., Vogel, B., et al. (2019). From era-interim to era5: the considerable impact of ecmwf's next-generation reanalysis on lagrangian transport simulations. *Atmospheric Chemistry and Physics*, 19(5):3097–3124.

Hua, W., Zhou, L., Nicholson, S. E., Chen, H., and Qin, M. (2019). Assessing reanalysis data for understanding rainfall climatology and variability over central equatorial africa. *Climate dynamics*, 53(1):651–669.

Huang, B., Banzon, V. F., Freeman, E., Lawrimore, J., Liu, W., Peterson, T. C., Smith, T. M., Thorne, P. W., Woodruff, S. D., and Zhang, H.-M. (2015). Extended reconstructed sea surface temperature version 4 (ersst. v4). part i: Upgrades and intercomparisons. *Journal of climate*, 28(3):911–930.

Hulme, M., Doherty, R., Ngara, T., New, M., and Lister, D. (2001). African climate change: 1900-2100. *Climate research*, 17(2):145–168.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models. *Journal of Geophysical Research: Atmospheres*, 113(D13).

Iizumi, T., Furuya, J., Shen, Z., Kim, W., Okada, M., Fujimori, S., Hasegawa, T., and Nishimori, M. (2017). Responses of crop yield growth to global temperature and socioeconomic changes. *Scientific reports*, 7(1):1–10.

Iizumi, T., Luo, J.-J., Challinor, A. J., Sakurai, G., Yokozawa, M., Sakuma, H., Brown, M. E., and Yamagata, T. (2014). Impacts of el niño southern oscillation on the global yields of major crops. *Nature communications*, 5(1):1–7.

Iizumi, T. and Ramankutty, N. (2016). Changes in yield variability of major crops for 1981–2010 explained by climate change. *Environmental research letters*, 11(3):034003.

Iizumi, T. and Sakai, T. (2020). The global dataset of historical yields for major crops 1981–2016. *Scientific data*, 7(1):1–7.

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., and Núñez-Riboni, I. (2013). Global ocean biogeochemistry model hamocc: Model architecture and performance as component of the mpi-earth system model in different cmip5 experimental realizations. *Journal of Advances in Modeling Earth Systems*, 5(2):287–315.

Indeje, M., Semazzi, F. H., and Ogallo, L. J. (2000). Enso signals in east african rainfall seasons. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 20(1):19–46.

Jayanthi, H. and Husak, G. (2013). A probabilistic approach to assess agricultural drought risk. *Background Paper Prepared for the Global Assessment Report on Disaster Risk Reduction*.

Jayanthi, H., Husak, G. J., Funk, C., Magadzire, T., Chavula, A., and Verdin, J. P. (2013). Modeling rain-fed maize vulnerability to droughts using the standardized precipitation index from satellite estimated rainfall—southern malawi case study. *International Journal of Disaster Risk Reduction*, 4:71–81.

Jayanthi, L., Sekar, J., Ameer Basha, S., and Parthasarathi, K. (2014). Influence of vermifertilizer on soil quality, yield and quality of chilli, capsicum annuum. *Online International Interdisciplinary Research Journal*, 4(Special Issue (March)):206–218.

Johnson, F. and Sharma, A. (2012). A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. *Water Resources Research*, 48(1).

Jolliffe, I. T. (2002). Graphical representation of data using principal components. *Principal component analysis*, pages 78–110.

Jones, P. D. and Moberg, A. (2003). Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *Journal of climate*, 16(2):206–223.

Jones, P. G. and Thornton, P. K. (2003). The potential impacts of climate change on maize production in africa and latin america in 2055. *Global environmental change*, 13(1):51–59.

Jungclaus, J., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and Von Storch, J. (2013). Characteristics of the ocean simulations in the max planck institute ocean model (mpiom) the ocean component of the mpi-earth system model. *Journal of Advances in Modeling Earth Systems*, 5(2):422–446.

Kaplan, A., Cane, M. A., Kushnir, Y., Clement, A. C., Blumenthal, M. B., and Rajagopalan, B. (1998). Analyses of global sea surface temperature 1856–1991. *Journal of Geophysical Research: Oceans*, 103(C9):18567–18589.

Kirtman, B., Power, S. B., Adedoyin, A. J., Boer, G. J., Bojariu, R., Camilloni, I., Doblas-Reyes, F., Fiore, A. M., Kimoto, M., Meehl, G., et al. (2013). Near-term climate change: projections and predictability.

Klutse, N. A. B., Abiodun, B. J., Hewitson, B. C., Gutowski, W. J., and Tadross, M. A. (2016). Evaluation of two gcms in simulating rainfall inter-annual variability over southern africa. *Theoretical and applied climatology*, 123(3-4):415–436.

Knorr, W. and Schnitzler, K.-G. (2006). Enhanced albedo feedback in north africa from possible combined vegetation and soil-formation processes. *Climate dynamics*, 26(1):55–63.

Knox, J., Daccache, A., Hess, T., and Haro, D. (2016). Meta-analysis of climate impacts and uncertainty on crop yields in europe. *Environmental Research Letters*, 11(11):113004.

Knox, J., Hess, T., Daccache, A., and Wheeler, T. (2012). Climate change impacts on crop productivity in africa and south asia. *Environmental research letters*, 7(3):034032.

Kotir, J. H. (2011). Climate change and variability in sub-saharan africa: a review of current and future trends and impacts on agriculture and food security. *Environment, Development and Sustainability*, 13(3):587–605.

Kristjanson, P., Neufeldt, H., Gassner, A., Mango, J., Kyazze, F. B., Desta, S., Sayula, G., Thiede, B., Förch, W., Thornton, P. K., et al. (2012). Are food insecure smallholder households making changes in their farming practices? evidence from east africa. *Food Security*, 4(3):381–397.

Kucharski, F., Bracco, A., Yoo, J., Tompkins, A., Feudale, L., Ruti, P., and Dell'Aquila, A. (2009). A gill–matsuno-type mechanism explains the tropical atlantic influence on african and indian monsoon rainfall. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 135(640):569–579.

Kurukulasuriya, P., Mendelsohn, R., Hassan, R., Benhin, J., Deressa, T., Diop, M., Eid, H. M., Fosu, K. Y., Gbetibouo, G., Jain, S., et al. (2006). Will african agriculture survive climate change? *The World Bank Economic Review*, 20(3):367–388.

Lamb, P. J. and Peppler, R. A. (1987). North atlantic oscillation: concept and an application. *Bulletin of the American Meteorological Society*, 68(10):1218–1225.

Lawrence, D. and Vandecar, K. (2015). Effects of tropical deforestation on climate and agriculture. *Nature climate change*, 5(1):27–36.

Leakey, A. D., Ainsworth, E. A., Bernacchi, C. J., Rogers, A., Long, S. P., and Ort, D. R. (2009). Elevated co2 effects on plant carbon, nitrogen, and water relations: six important lessons from face. *Journal of experimental botany*, 60(10):2859–2876.

Leander, R. and Buishand, T. A. (2007). Resampling of regional climate model output for the simulation of extreme river flows. *Journal of Hydrology*, 332(3-4):487–496.

Legesse, G. (2010). Agricultural drought assessment using remote sensing and gis techniques. *Addis Ababa University, Ethiopia*.

Lenderink, G., Buishand, A., and Deursen, W. v. (2007). Estimates of future discharges of the river rhine using two scenario methodologies: direct versus delta approach. *Hydrology and Earth System Sciences*, 11(3):1145–1159.

Lin, S.-J. and Rood, R. B. (1996). Multidimensional flux-form semi-lagrangian transport schemes. *Monthly Weather Review*, 124(9):2046–2070.

Lindesay, J. (1988). South african rainfall, the southern oscillation and a southern hemisphere semi-annual cycle. *Journal of climatology*, 8(1):17–30.

Lobell, D., Bala, G., and Duffy, P. (2006). Biogeophysical impacts of cropland management changes on climate. *Geophysical Research Letters*, 33(6).

Lobell, D. B. and Asseng, S. (2017). Comparing estimates of climate change impacts from process-based and statistical crop models. *Environmental Research Letters*, 12(1):015001.

Lobell, D. B., Bänziger, M., Magorokosho, C., and Vivek, B. (2011). Nonlinear heat effects on african maize as evidenced by historical yield trials. *Nature climate change*, 1(1):42–45.

Lobell, D. B. and Burke, M. (2009). *Climate change and food security: adapting agriculture to a warmer world*, volume 37. Springer Science & Business Media.

Lobell, D. B. and Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and forest meteorology*, 150(11):1443–1452.

Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., and Naylor, R. L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science*, 319(5863):607–610.

Lobell, D. B. and Field, C. B. (2007). Global scale climate–crop yield relationships and the impacts of recent warming. *Environmental research letters*, 2(1):014002.

Losada, T., Rodríguez-Fonseca, B., Janicot, S., Gervois, S., Chauvin, F., and Ruti, P. (2010). A multi-model approach to the atlantic equatorial mode: impact on the west african monsoon. *Climate Dynamics*, 35(1):29–43.

Losada, T., Rodriguez-Fonseca, B., Mohino, E., Bader, J., Janicot, S., and Mechoso, C. R. (2012). Tropical sst and sahel rainfall: A non-stationary relationship. *Geophysical Research Letters*, 39(12).

Lott, F. (1999). Alleviation of stationary biases in a gcm through a mountain drag parameterization scheme and a simple representation of mountain lift forces. *Monthly weather review*, 127(5):788–801.

Lu, J. and Delworth, T. L. (2005). Oceanic forcing of the late 20th century sahel drought. *Geophysical Research Letters*, 32(22).

Lynd, L. R. and Woods, J. (2011). A new hope for africa: Bioenergy could help bring food security to the world's poorest continent. *Nature*, 474(7352):S20–S20.

Malhi, Y. and Wright, J. (2004). Spatial patterns and recent trends in the climate of tropical rainforest regions. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1443):311–329.

Malsy, M., aus der Beek, T., and Flörke, M. (2015). Evaluation of large-scale precipitation data sets for water resources modelling in central asia. *Environmental Earth Sciences*, 73(2):787–799.

Mansouri-Far, C., Sanavy, S. A. M. M., and Saberali, S. F. (2010). Maize yield response to deficit irrigation during low-sensitive growth stages and nitrogen rate under semi-arid climatic conditions. *Agricultural Water Management*, 97(1):12–22.

Maraun, D. (2016). Bias correcting climate change simulations-a critical review. *Current Climate Change Reports*, 2(4):211–220.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M., Hall, A., et al. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7(11):764–773.

Maraun, D., Widmann, M., and Gutiérrez, J. M. (2019). Statistical downscaling skill under present climate conditions: A synthesis of the value perfect predictor experiment. *International Journal of Climatology*, 39(9):3692–3703.

Martin, R. V., Washington, R., and Downing, T. E. (2000). Seasonal maize forecasting for south africa and zimbabwe derived from an agroclimatological model. *Journal of Applied Meteorology*, 39(9):1473–1479.

Martinez-Artigas, J., Lemus-Canovas, M., and Lopez-Bustins, J. A. (2021). Precipitation in peninsular spain: Influence of teleconnection indices and spatial regionalisation. *International Journal of Climatology*, 41:E1320–E1335.

Masasi, B. and Ng'ombe, J. N. (2019). Does a market systems approach revitalize smallholder irrigation schemes? evidence from zimbabwe.

Mason, S. and Tyson, P. (1992). The modulation of sea surface temperature and rainfall associations over southern africa with solar activity and the quasi-biennial oscillation. *Journal of Geophysical Research: Atmospheres*, 97(D5):5847–5856.

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., et al. (2018). Global warming of 1.5 c. *An IPCC Special Report on the impacts of global warming of*, 1:1–9.

Masupha, T. E. and Moeletsi, M. E. (2018). Analysis of potential future droughts limiting maize production, in the luvuvhu river catchment area, south africa. *Physics and Chemistry of the Earth, Parts A/B/C*, 105:44–51.

McGregor, J. L. (2015). Recent developments in variable-resolution global climate modelling. *Climatic Change*, 129(3):369–380.

McNally, A., Husak, G. J., Brown, M., Carroll, M., Funk, C., Yatheendradas, S., Arsenault, K., Peters-Lidard, C., and Verdin, J. P. (2015). Calculating crop water requirement satisfaction in the west africa sahel with remotely sensed soil moisture. *Journal of Hydrometeorology*, 16(1):295–305.

McVicar, T. R., Van Niel, T. G., Li, L., Hutchinson, M. F., Mu, X., and Liu, Z. (2007). Spatially distributing monthly reference evapotranspiration and pan evaporation considering topographic influences. *Journal of hydrology*, 338(3-4):196–220.

Megevand, C. (2013). *Dynamiques de déforestation dans le basin du Congo: Réconcilier la croissance économique et la protection de la forêt*. World Bank Publications.

Merchant, C. J., Embury, O., Rayner, N. A., Berry, D. I., Corlett, G. K., Lean, K., Veal, K. L., Kent, E. C., Llewellyn-Jones, D. T., Remedios, J. J., et al. (2012). A 20 year independent record of sea surface temperature for climate from along-track scanning radiometers. *Journal of Geophysical Research: Oceans*, 117(C12).

Milesi, J., Toteu, S., Deschamps, Y., Feybesse, J., Lerouge, C., Cocherie, A., Penaye, J., Tchameni, R., Moloto-A-Kenguemba, G., Kampunzu, H., et al. (2006). An overview of the geology and major ore deposits of central africa: Explanatory note for the 1: 4,000,000 map "geology and major ore deposits of central africa". *Journal of African Earth Sciences*, 44(4-5):571–595.

Mishra, A. K. and Singh, V. P. (2010). A review of drought concepts. *Journal of hydrology*, 391(1-2):202–216.

Mohino, E., Janicot, S., and Bader, J. (2011). Sahel rainfall and decadal to multi-decadal sea surface temperature variability. *Climate dynamics*, 37(3-4):419–440.

Moore, S. K., Trainer, V. L., Mantua, N. J., Parker, M. S., Laws, E. A., Backer, L. C., and Fleming, L. E. (2008). Impacts of climate variability and future climate change on harmful algal blooms and human health. In *Environmental Health*, volume 7, pages 1–12. BioMed Central.

Mulenga, B. P., Wineman, A., and Sitko, N. J. (2017). Climate trends and farmers' perceptions of climate change in zambia. *Environmental management*, 59(2):291–306.

Müller, C., Cramer, W., Hare, W. L., and Lotze-Campen, H. (2011). Climate change risks for african agriculture. *Proceedings of the National Academy of Sciences*, 108(11):4313–4315.

Naidu, P. D., Ganeshram, R., Bollasina, M. A., Panmei, C., Nürnberg, D., and Donges, J. F. (2020). Coherent response of the indian monsoon rainfall to atlantic multi-decadal variability over the last 2000 years. *Scientific reports*, 10(1):1–11.

Nakicenovic, N., Alcamo, J., Davis, G., Vries, B. d., Fenhann, J., Gaffin, S., Gregory, K., Grubler, A., Jung, T. Y., Kram, T., et al. (2000). Special report on emissions scenarios.

Nelson, G. C., Rosegrant, M. W., Koo, J., Robertson, R., Sulser, T., Zhu, T., Ringler, C., Msangi, S., Palazzo, A., Batka, M., et al. (2009). *Climate change: Impact on agriculture and costs of adaptation*, volume 21. Intl Food Policy Res Inst.

Nesbitt, S. W., Cifelli, R., and Rutledge, S. A. (2006). Storm morphology and rainfall characteristics of trmm precipitation features. *Monthly Weather Review*, 134(10):2702–2721.

New, M., Hewitson, B., Stephenson, D. B., Tsiga, A., Kruger, A., Manhique, A., Gomez, B., Coelho, C. A., Masisi, D. N., Kululanga, E., et al. (2006). Evidence of trends in daily climate extremes over southern and west africa. *Journal of Geophysical Research: Atmospheres*, 111(D14).

New, M., Hulme, M., and Jones, P. (2000). Representing twentieth-century space–time climate variability. part ii: Development of 1901–96 monthly grids of terrestrial surface climate. *Journal of climate*, 13(13):2217–2238.

Ngoma, H., Hamududu, B., Hangoma, P., Samboko, P., Hichaambwa, M., and Kabaghe, C. (2019). Irrigation development for climate resilience in zambia: The known knowns and known unknowns. Technical report.

Ng'ongolo, H. K., Smyshlyaev, S. P., et al. (2010). The statistical prediction of east african rainfalls using quasi-biennial oscillation phases information. *Natural Science*, 2(12):1407.

Nicholson, S. E. (2017). Climate and climatic variability of rainfall over eastern africa. *Reviews of Geophysics*, 55(3):590–635.

Nicholson, S. E. (2018). Climate of the sahel and west africa. In *Oxford Research Encyclopedia of Climate Science*.

Nicholson, S. E. (2019). A review of climate dynamics and climate variability in eastern africa. *The limnology, climatology and paleoclimatology of the East African lakes*, pages 25–56.

Nicholson, S. E. and Dezfuli, A. K. (2013). The relationship of rainfall variability in western equatorial africa to the tropical oceans and atmospheric circulation. part i: The boreal spring. *Journal of climate*, 26(1):45–65.

Nicholson, S. E. and Grist, J. P. (2003). The seasonal evolution of the atmospheric circulation over west africa and equatorial africa. *Journal of climate*, 16(7):1013–1030.

Nicholson, S. E. and Selato, J. (2000). The influence of la nina on african rainfall. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 20(14):1761–1776.

Nikulin, G., Jones, C., Giorgi, F., Asrar, G., Büchner, M., Cerezo-Mota, R., Christensen, O. B., Déqué, M., Fernandez, J., Hänsler, A., et al. (2012). Precipitation climatology in an ensemble of cordex-africa regional climate simulations. *Journal of Climate*, 25(18):6057–6078.

Nnamchi, H. C., Li, J., and Anyadike, R. N. (2011). Does a dipole mode really exist in the south atlantic ocean? *Journal of Geophysical Research: Atmospheres*, 116(D15).

Nogherotto, R., Coppola, E., Giorgi, F., and Mariotti, L. (2013). Impact of congo basin deforestation on the african monsoon. *Atmospheric Science Letters*, 14(1):45–51.

OECD/FAO (2016). Agriculture in sub-saharan africa: Prospects and challenges for the next decade. *OECD-FAO Agricultural Outlook 2016-2025*.

Ongoma, V. and Chen, H. (2017). Temporal and spatial variability of temperature and precipitation over east africa from 1951 to 2010. *Meteorology and Atmospheric Physics*, 129(2):131–144.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P., et al. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Ipcc.

Paeth, H., Born, K., Girmes, R., Podzun, R., and Jacob, D. (2009). Regional climate change in tropical and northern africa due to greenhouse forcing and land use changes. *Journal of Climate*, 22(1):114–132.

Paeth, H., Born, K., Podzun, R., and Jacob, D. (2005). Regional dynamical downscaling over west africa: model evaluation and comparison of wet and dry years. *Meteorologische Zeitschrift*, 14(3):349–367.

Paeth, H., Capo-Chichi, A., and Endlicher, W. (2008). Climate change and food security in tropical west africa—a dynamic-statistical modelling approach. *Erdkunde*, pages 101–115.

Paeth, H. and Hense, A. (2003). Seasonal forecast of sub-sahelian rainfall using cross validated model output statistics. *Meteorologische zeitschrift*, pages 157–173.

Panitz, S., Salzmann, U., Risebrobakken, B., De Schepper, S., and Pound, M. J. (2016). Climate variability and long-term expansion of peatlands in arctic norway during the late pliocene (odp site 642, norwegian sea). *Climate of the Past*, 12(4):1043–1060.

Panofsky, H. A. and Brier, G. W. (1968). *Some applications of statistics to meteorology*. Earth and Mineral Sciences Continuing Education, College of Earth and ....

Parker, D., Jones, P., Folland, C., and Bevan, A. (1994). Interdecadal changes of surface temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, 99(D7):14373–14399.

Parkes, B., Sultan, B., and Ciais, P. (2018). The impact of future climate change and potential adaptation methods on maize yields in west africa. *Climatic Change*, 151(2):205–217.

Pedgley, D. (1967). Weather in the mountains. *Weather*, 22(7):266–275.

Piani, C., Weedon, G., Best, M., Gomes, S., Viterbo, P., Hagemann, S., and Haerter, J. (2010). Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of hydrology*, 395(3-4):199–215.

Popke, D., Stevens, B., and Voigt, A. (2013). Climate and climate change in a radiative-convective equilibrium version of echam6. *Journal of Advances in Modeling Earth Systems*, 5(1):1–14.

Porter, J. R., Xie, L., Challinor, A. J., Cochrane, K., Howden, S. M., Iqbal, M. M., Lobell, D. B., and Travasso, M. I. (2014). Food security and food production systems.

Prasad, P., Staggenborg, S., and Ristic, Z. (2008). Impacts of drought and/or heat stress on physiological, developmental, growth, and yield processes of crop plants. *Response of crops to limited water: Understanding and modeling water stress effects on plant growth processes*, 1:301–355.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., et al. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of geophysics*, 53(2):323–361.

Rafi, Z. and Ahmad, R. (2005). Wheat crop model based on water balance for agrometeorological crop monitoring. *Pakistan Journal of Meteorology Vol*, 2(3).

Ramankutty, N., Evan, A. T., Monfreda, C., and Foley, J. A. (2008). Farming the planet: 1. geographic distribution of global agricultural lands in the year 2000. *Global biogeochemical cycles*, 22(1).

Ravallion, M. (2012). Benchmarking global poverty reduction. *World Bank Policy Research Working Paper*, (6205).

Ray, D. K., Gerber, J. S., MacDonald, G. K., and West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nature communications*, 6(1):1–9.

Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., and Foley, J. A. (2012). Recent patterns of crop yield growth and stagnation. *Nature communications*, 3(1):1–7.

Reason, C. and Jagadheesha, D. (2005). A model investigation of recent enso impacts over southern africa. *Meteorology and Atmospheric Physics*, 89(1):181–205.

Reick, C., Raddatz, T., Brovkin, V., and Gayler, V. (2013). Representation of natural and anthropogenic land cover change in mpi-esm. *Journal of Advances in Modeling Earth Systems*, 5(3):459–482.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W. (2002). An improved in situ and satellite sst analysis for climate. *Journal of climate*, 15(13):1609–1625.

Riahi, K., Grübler, A., and Nakicenovic, N. (2007). Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technological Forecasting and Social Change*, 74(7):887–935.

Ridolfi, L., D'Odorico, P., Laio, F., Tamea, S., and Rodriguez-Iturbe, I. (2008). Coupled stochastic dynamics of water table and soil moisture in bare soil conditions. *Water resources research*, 44(1).

Rim, C.-S. (2009). The effects of urbanization, geographical and topographical conditions on reference evapotranspiration. *Climatic Change*, 97(3):483–514.

Rizou, D., Flocas, H. A., Hatzaki, M., and Bartzokas, A. (2018). A statistical investigation of the impact of the indian monsoon on the eastern mediterranean circulation. *Atmosphere*, 9(3):90.

Roberts, M. J., Braun, N. O., Sinclair, T. R., Lobell, D. B., and Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, 12(9):095010.

Roberts, M. J., Schlenker, W., and Eyer, J. (2013). Agronomic weather measures in econometric models of crop yield with implications for climate change. *American Journal of Agricultural Economics*, 95(2):236–243.

Rojas, R., Feyen, L., Dosio, A., and Bavera, D. (2011). Improving pan-european hydrological simulation of extreme events through statistical bias correction of rcm-driven climate simulations. *Hydrology and Earth System Sciences*, 15(8):2599–2620.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., et al. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, 111(9):3268–3273.

Roudier, P., Sultan, B., Quirion, P., and Berg, A. (2011). The impact of future climate change on west african crop yields: What does the recent literature say? *Global environmental change*, 21(3):1073–1083.

Rummukainen, M. (2010). State-of-the-art with regional climate models. *Wiley Interdisciplinary Reviews: Climate Change*, 1(1):82–96.

Rummukainen, M. (2016). Added value in regional climate modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 7(1):145–159.

Rurinda, J., Mapfumo, P., Van Wijk, M., Mtambanengwe, F., Rufino, M. C., Chikowo, R., and Giller, K. E. (2014). Sources of vulnerability to a variable and changing climate among smallholder households in zimbabwe: A participatory analysis. *Climate Risk Management*, 3:65–78.

Sacks, W. J., Deryng, D., Foley, J. A., and Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global ecology and biogeography*, 19(5):607–620.

Saeed, F., Bethke, I., Fischer, E., Legutke, S., Shiogama, H., Stone, D. A., and Schleussner, C.-F. (2018). Robust changes in tropical rainy season length at 1.5 c and 2 c. *Environmental Research Letters*, 13(6):064024.

Saji, N., Goswami, B., Vinayachandran, P., and Yamagata, T. (1999). A dipole mode in the tropical indian ocean. *Nature*, 401(6751):360–363.

Salami, A., Kamara, A. B., and Brixiova, Z. (2010). *Smallholder agriculture in East Africa: Trends, constraints and opportunities*. African Development Bank Tunis.

Samanta, S., Pal, D. K., Lohar, D., and Pal, B. (2012). Interpolation of climate variables and temperature modeling. *Theoretical and Applied Climatology*, 107(1):35–45.

Sangelantoni, L., Russo, A., and Gennaretti, F. (2019). Impact of bias correction and downscaling through quantile mapping on simulated climate change signal: a case study over central italy. *Theoretical and Applied Climatology*, 135(1):725–740.

Schlenker, W. and Lobell, D. B. (2010). Robust negative impacts of climate change on african agriculture. *Environmental Research Letters*, 5(1):014010.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of sciences*, 106(37):15594–15598.

Schmidt, G. A., Ruedy, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., Bauer, S., Cairns, B., Canuto, V., Cheng, Y., et al. (2006). Present-day atmospheric simulations using giss modele: Comparison to in situ, satellite, and reanalysis data. *Journal of Climate*, 19(2):153–192.

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

Schubert, J. J., Stevens, B., and Crueger, T. (2013). Madden-julian oscillation as simulated by the mpi earth system model: Over the last and into the next millennium. *Journal of Advances in Modeling Earth Systems*, 5(1):71–84.

Schwan, S. and Yu, X. (2018). Social protection as a strategy to address climate-induced migration. *International Journal of Climate Change Strategies and Management*.

Senay, G. B. and Verdin, J. (2003). Characterization of yield reduction in ethiopia using a gis-based crop water balance model. *Canadian Journal of Remote Sensing*, 29(6):687–692.

Sennikovs, J. and Bethers, U. (2009). Statistical downscaling method of regional climate model results for hydrological modelling. In *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, edited by: Anderssen, RS, Braddock, RD, and Newham, LTH*, pages 3962–3968. Citeseer.

Shanahan, T. M. (2018). Quaternary climate variation in west africa. In *Oxford Research Encyclopedia of Climate Science*.

Sharif, B., Makowski, D., Plauborg, F., and Olesen, J. E. (2017). Comparison of regression techniques to predict response of oilseed rape yield to variation in climatic conditions in denmark. *European Journal of Agronomy*, 82:11–20.

Shi, H., Li, T., and Wei, J. (2017). Evaluation of the gridded cru ts precipitation dataset with the point raingauge records over the three-river headwaters region. *Journal of Hydrology*, 548:322–332.

Shi, W. and Tao, F. (2014). Vulnerability of african maize yield to climate change and variability during 1961–2010. *Food Security*, 6(4):471–481.

Shongwe, M. E., van Oldenborgh, G. J., van den Hurk, B., and van Aalst, M. (2011). Projected changes in mean and extreme precipitation in africa under global warming. part ii: East africa. *Journal of climate*, 24(14):3718–3733.

Sillmann, J., Kharin, V. V., Zwiers, F., Zhang, X., and Bronaugh, D. (2013). Climate extremes indices in the cmip5 multimodel ensemble: Part 2. future climate projections. *Journal of Geophysical Research: Atmospheres*, 118(6):2473–2493.

Simmons, A. J. and Burridge, D. M. (1981). An energy and angular-momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. *Monthly Weather Review*, 109(4):758–766.

Smith, P. C., Heinrich, G., Suklitsch, M., Gobiet, A., Stoffel, M., and Fuhrer, J. (2014). Station-scale bias correction and uncertainty analysis for the estimation of irrigation water requirements in the swiss rhone catchment under climate change. *Climatic Change*, 127(3):521–534.

Sneyers, R. et al. (1991). *On the statistical analysis of series of observations.* Number 143.

Sonkoué, D., Monkam, D., Fotso-Nguemo, T. C., Yepdo, Z. D., and Vondou, D. A. (2019). Evaluation and projected changes in daily rainfall characteristics over central africa based on a multi-model ensemble mean of cmip5 simulations. *Theoretical and Applied Climatology*, 137(3):2167–2186.

Srivastava, A., Pradhan, M., Goswami, B., and Rao, S. A. (2019a). Regime shift of indian summer monsoon rainfall to a persistent arid state: external forcing versus internal variability. *Meteorology and Atmospheric Physics*, 131(2):211–224.

Srivastava, A. K., Mboh, C. M., Gaiser, T., Kuhn, A., Ermias, E., and Ewert, F. (2019b). Effect of mineral fertilizer on rain water and radiation use efficiencies for maize yield and stover biomass productivity in ethiopia. *Agricultural Systems*, 168:88–100.

Srivastava, G., Chakraborty, A., and Nanjundiah, R. S. (2019c). Multidecadal see-saw of the impact of enso on indian and west african summer monsoon rainfall. *Climate dynamics*, 52(11):6633–6649.

Staver, A. C., Archibald, S., and Levin, S. (2011). Tree cover in sub-saharan africa: rainfall and fire constrain forest and savanna as alternative stable states. *Ecology*, 92(5):1063–1072.

Stevens, B., Bony, S., and Webb, M. (2012). Clouds on-off klimate intercomparison experiment (cookie).

Sultan, B. and Gaetani, M. (2016). Agriculture in west africa in the twenty-first century: climate change and impacts scenarios, and potential for adaptation. *Frontiers in Plant Science*, 7:1262.

Sultan, B., Guan, K., Kouressy, M., Biasutti, M., Piani, C., Hammer, G., McLean, G., and Lobell, D. B. (2014). Robust features of future climate change impacts on sorghum yields in west africa. *Environmental Research Letters*, 9(10):104006.

Sundqvist, H., Berge, E., and KRISTJANSSON, J. (1989). Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Monthly Weather Review*, 117(8):1641–1657.

Tamoffo, A. T., Moufouma-Okia, W., Dosio, A., James, R., Pokam, W. M., Vondou, D. A., Fotso-Nguemo, T. C., Guenang, G. M., Kamsu-Tamo, P. H., Nikulin, G., et al. (2019a). Process-oriented assessment of rca4 regional climate model projections over the congo basin under 1.5\^{\circ}{\text {C}}

$$1.5cand$$

2\^{\circ}{\text {C}}

$2cglobalwarminglevels : influenceofregionalmoisturefluxes.ClimateDynamics, 53(3) : 1911−−19$

Tamoffo, A. T., Vondou, D. A., Pokam, W. M., Haensler, A., Yepdo, Z. D., Fotso-Nguemo, T. C., Tchotchou, L. A. D., and Nouayou, R. (2019b). Daily characteristics of central african rainfall in the remo model. *Theoretical and Applied Climatology*, 137(3):2351–2368.

Tao, F., Yokozawa, M., Liu, J., and Zhang, Z. (2008). Climate–crop yield relationships at provincial scales in china and the impacts of recent climate trends. *Climate Research*, 38(1):83–94.

Tao, F. and Zhang, Z. (2011). Impacts of climate change as a function of global mean temperature: maize productivity and water use in china. *Climatic Change*, 105(3):409–432.

Tao, F., Zhang, Z., Xiao, D., Zhang, S., Rötter, R. P., Shi, W., Liu, Y., Wang, M., Liu, F., and Zhang, H. (2014). Responses of wheat growth and yield to climate change in different climate zones of china, 1981–2009. *Agricultural and Forest Meteorology*, 189:91–104.

Tarnavsky, E., Chavez, E., and Boogaard, H. (2018). Agro-meteorological risks to maize production in tanzania: Sensitivity of an adapted water requirements satisfaction index (wrsi) model to rainfall. *International Journal of Applied Earth Observation and Geoinformation*, 73:77–87.

Taylor, C. M., Birch, C. E., Parker, D. J., Dixon, N., Guichard, F., Nikulin, G., and Lister, G. M. (2013). Modeling soil moisture-precipitation feedback in the sahel: Impor-

tance of spatial scale versus convective parameterization. *Geophysical Research Letters*, 40(23):6213–6218.

Taylor, K., Stouffer, R., Meehl, G., Taylor, K., Stouffer, R., and Meehl, G. (2012). An overview of cmip5 and the experiment design, b. am. meteorol. soc., 93, 485–498.

Tebaldi, C. and Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical transactions of the royal society A: mathematical, physical and engineering sciences*, 365(1857):2053–2075.

Teichmann, C., Eggert, B., Elizalde, A., Haensler, A., Jacob, D., Kumar, P., Moseley, C., Pfeifer, S., Rechid, D., Remedio, A. R., et al. (2013). How does a regional climate model modify the projected climate change signal of the driving gcm: a study over different cordex regions using remo. *Atmosphere*, 4(2):214–236.

Tesfaye, K., Gbegbelegbe, S., Cairns, J. E., Shiferaw, B., Prasanna, B. M., Sonder, K., Boote, K., Makumbi, D., and Robertson, R. (2015). Maize systems under climate change in sub-saharan africa. *International Journal of Climate Change Strategies and Management*.

Teutschbein, C. and Seibert, J. (2010). Regional climate models for hydrological impact studies at the catchment scale: a review of recent modeling strategies. *Geography Compass*, 4(7):834–860.

Teutschbein, C. and Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of hydrology*, 456:12–29.

Themeßl, M. J., Gobiet, A., and Heinrich, G. (2012). Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, 112(2):449–468.

Thornton, P. K., Jones, P. G., Ericksen, P. J., and Challinor, A. J. (2011). Agriculture and food systems in sub-saharan africa in a 4 c+ world. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1934):117–136.

Thornton, P. K., van de Steeg, J., Notenbaert, A., and Herrero, M. (2009). The impacts of climate change on livestock and livestock systems in developing countries: A review of what we know and what we need to know. *Agricultural systems*, 101(3):113–127.

Tian, X. and Yu, X. (2019). Crop yield gap and yield convergence in african countries. *Food Security*, 11(6):1305–1319.

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly weather review*, 117(8):1779–1800.

Tingem, M. and Rivington, M. (2009). Adaptation for crop agriculture to climate change in cameroon: turning on the heat. *Mitigation and Adaptation Strategies for Global Change*, 14(2):153–168.

Trenberth, K. E., Jones, P. D., Ambenje, P., Bojariu, R., Easterling, D., Klein Tank, A., Parker, D., Rahimzadeh, F., Renwick, J. A., Rusticucci, M., et al. (2007). Observations. surface and atmospheric climate change. chapter 3.

Turner, W. A. (2020). *An Improved Climatological Forecast Method for Projecting End-Of-Season Water Requirement Satisfaction Index (WRSI)*. University of California, Santa Barbara.

Urban, D. W., Roberts, M. J., Schlenker, W., and Lobell, D. B. (2015). The effects of extremely wet planting conditions on maize and soybean yields. *Climatic Change*, 130(2):247–260.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., et al. (2011). The representative concentration pathways: an overview. *Climatic change*, 109(1):5–31.

Verdin, J. and Klaver, R. (2002). Grid-cell-based crop water accounting for the famine early warning system. *Hydrological Processes*, 16(8):1617–1630.

Vicente-Serrano, S. M., Van der Schrier, G., Beguería, S., Azorin-Molina, C., and Lopez-Moreno, J.-I. (2015). Contribution of precipitation and reference evapotranspiration to drought indices under different climates. *Journal of Hydrology*, 526:42–54.

Vizy, E. K. and Cook, K. H. (2012). Mid-twenty-first-century changes in extreme events over northern and tropical africa. *Journal of Climate*, 25(17):5748–5767.

Von Braun, J. (2007). *The world food situtation: new driving forces and required actions*. Intl Food Policy Res Inst.

von Storch, H. (1999). Spatial patterns: Eofs and cca. In *Analysis of climate variability*, pages 231–263. Springer.

Vondou, D. A. and Haensler, A. (2017). Evaluation of simulations with the regional climate model remo over central africa and the effect of increased spatial resolution. *International Journal of Climatology*, 37:741–760.

Waha, K., Müller, C., Bondeau, A., Dietrich, J. P., Kurukulasuriya, P., Heinke, J., and Lotze-Campen, H. (2013). Adaptation to climate change through the choice of cropping system and sowing date in sub-saharan africa. *Global Environmental Change*, 23(1):130–143.

Wainwright, C. M., Finney, D. L., Kilavi, M., Black, E., and Marsham, J. H. (2021). Extreme rainfall in east africa, october 2019–january 2020 and context under future climate change. *Weather*, 76(1):26–31.

Wilks, D. S. (2006). Comparison of ensemble-mos methods in the lorenz'96 setting. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, 13(3):243–256.

Willcock, S., Phillips, O. L., Platts, P. J., Swetnam, R. D., Balmford, A., Burgess, N. D., Ahrends, A., Bayliss, J., Doggart, N., Doody, K., et al. (2016). Land cover change and carbon emissions over 100 years in an a frican biodiversity hotspot. *Global Change Biology*, 22(8):2787–2800.

Winch, T. (2006). Section 1 the principles and practices used in agriculture and horticulture. *Growing Food: A Guide to Food Production*, pages 1–103.

Wise, M., Calvin, K., Thomson, A., Clarke, L., Bond-Lamberty, B., Sands, R., Smith, S. J., Janetos, A., and Edmonds, J. (2009). Implications of limiting co2 concentrations for land use and energy. *Science*, 324(5931):1183–1186.

Wolff, C., Haug, G. H., Timmermann, A., Damsté, J. S. S., Brauer, A., Sigman, D. M., Cane, M. A., and Verschuren, D. (2011). Reduced interannual rainfall variability in east africa during the last ice age. *Science*, 333(6043):743–747.

Wolter, K. (1987). The southern oscillation in surface circulation and climate over the tropical atlantic, eastern pacific, and indian oceans as captured by cluster analysis. *Journal of Applied Meteorology and Climatology*, 26(4):540–558.

Worou, K., Goosse, H., Fichefet, T., Guichard, F., and Diakhate, M. (2020). Interannual variability of rainfall in the guinean coast region and its links with sea surface temperature changes over the twentieth century for the different seasons. *Climate Dynamics*, 55:449–470.

Xu, C.-y., Gong, L., Jiang, T., Chen, D., and Singh, V. (2006). Analysis of spatial distribution and temporal trend of reference evapotranspiration and pan evaporation in changjiang (yangtze river) catchment. *Journal of hydrology*, 327(1-2):81–93.

Yuan, D., Hu, X., Xu, P., Zhao, X., Masumoto, Y., and Han, W. (2018). The iod-enso precursory teleconnection over the tropical indo-pacific ocean: Dynamics and long-term trends under global warming. *Journal of Oceanology and Limnology*, 36(1):4–19.

Yuan, Y., Li, C., and Yang, S. (2014). Decadal anomalies of winter precipitation over southern china in association with el niño and la niña. *Journal of Meteorological Research*, 28(1):91–110.

Zhai, R., Tao, F., Lall, U., and Elliott, J. (2021). Africa would need to import more maize in the future even under 1.5 c warming scenario. *Earth's Future*, 9(1):e2020EF001574.

Zhang, H., Loáiciga, H. A., Ha, D., and Du, Q. (2020). Spatial and temporal downscaling of trmm precipitation with novel algorithms. *Journal of Hydrometeorology*, 21(6):1259–1278.

Acknowledgments

This study was carried out at the Julius -Maximilians -University of Würzburg at the Chair of Physical Geography within the Institute of Geography and Geology . The project was initially funded by the Norwegian Agency for Development (NORAD) under the GrowNut Project (QZA-0484, COG-13/0002) in collaboration with the University of Kinshasa, the University of Kwazulu-Natal and the University of Bergen. The fund was completed by the Julius Maximilians University of Würzburg. I also received additional financial support from the Excellence Scholarship Program BEBUC. First, I would like to thank my supervisor Prof. Dr. Heiko Paeth for accepting me to conduct this research in his research group. I also thank him for his guidance, coordination and inspiration throughout the development of this research. I was particularly touched by his willingness to search for additional funds for me and his personal support in helping my wife to come to Germany. It was helpful for my emotional stability. I would also like to thank Prof. Dr. Barbara Sponholz for accepting to correct this thesis and making relevant suggestions. My special thanks go to Dr. Felix Pollinger for the technical support, especially in computer programming. Dr. Felix was always available to help. I would like to express my special thanks to the entire working group, especially to Daniel Abel, Luzia Keupp, Daniel Schönbein, Miriam Baumann, Christian Hartmann, Chibuike Ibebuchi, Alphonse Karama, Katrin Ziegler, Doro Schill, Praveen Rai and Mengjie Warmuth for their advice and provision of a friendly working atmosphere. Additionally, I would like to show gratitude to the GrowNut team, especially to Prof. Ali Mapatano, Prof. Pierre Z Akilimali, Prof. Paulin Mutombo, Prof. Marie-Claire Muyer, Prof. Anne Hatløy and Prof. Christiane Horwood for their endless support and encouragement. I could not have made it without their Belief in me. Finally, I deeply thank my wife, Dorcas Mulumba, for her endless encouragement at all times, particularly during the last phase of my thesis, which coincided with the birth of our blessed first child, Daniel. I also extend my thanks to my parents, Henry and Alphonsine, to whom I dedicate this work. Thank you for your support, Mum and Dad.