Towards Seasonal Forecasting of Maize Yield in eastern Africa:

Skill in the Forecast Model Chain as a Basis for Agricultural Climate Services

Geoffrey Evans Owino Ogutu

Towards Seasonal Forecasting of Maize Yield in eastern Africa
Geoffrey Evans Owino Ogutu

Propositions

1. The potential value of seasonal climate prediction in crop forecasting in East Africa is either unknown or underutilized (this thesis)

2. The use of many verification metrics to evaluate a forecast attribute, leads to weak conclusions on predictability (this thesis)

3. The obligation to use community based research tools such as statistical analysis packages, cripples creativity.

4. Early warning systems and related response systems should have a shared management to ensure timely actions and interventions.

5. Developing climate change adaptation plans based on local experiences are as effective as plans informed by science.

6. To alleviate food insecurity in developing countries, we need stronger scientific advocacy on modern technologies such as GMOs.

7. Formal training in leadership seldom transforms one into a leader.

8. Multidisciplinary PhD research limits development of expertise.

Propositions belonging to the thesis, entitled

Towards seasonal forecasting of maize yield in eastern Africa: Skill in the forecast model chain as a basis for agricultural climate services

Geoffrey Evans Owino Ogutu
Wageningen, 14th September 2020

Towards Seasonal Forecasting of Maize Yield in eastern Africa: Skill in the Forecast Model Chain as a Basis for Agricultural Climate Services

Geoffrey Evans Owino Ogutu

Thesis committee

Promoter

Dr R.W.A. Hutjes
Associate professor, Water Systems and Global Change Group,
Wageningen University & Research

Co-promoters

Dr Iwan Supit
Researcher, Water Systems and Global Change Group,
Wageningen University & Research

Dr Philip Omondi Amingo
Senior Researcher,
IGAD Climate Prediction and Applications Centre (ICPAC), Nairobi, Kenya

Other members
Prof. Dr Niels Anten, Wageningen University & Research
Prof. Dr B.J.J.M. van den Hurk, Vrije Universiteit Amsterdam
Prof. Dr W.A. Landman, University of Pretoria, Republic of South Africa
Prof. Dr Andy Challinor, University of Leeds, United Kingdom

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE)

Towards Seasonal Forecasting of Maize Yield in eastern Africa: Skill in the Forecast Model Chain as a Basis for Agricultural Climate Services

Geoffrey Evans Owino Ogutu

Thesis

Submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University & Research
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 14 September 2020
at 1:30 p.m. in the Aula

Geoffrey Evans Owino Ogutu

Towards Seasonal Forecasting of Maize Yield in eastern Africa: Skill in the Forecast Model Chain as a Basis for Agricultural Climate Services, 229 pages

PhD thesis, Wageningen University & Research, NL (2020)
With references, with summaries in English and Dutch

ISBN 978-94-6395-500-3
DOI https://doi.org/10.18174/529199

Abstract

Climate variability is an important driver for regionally anomalous production levels of especially rainfed crops, with implication for food security of subsistence farmers and economic performance for market oriented agriculturalists. In large parts of the tropics, modern seasonal ensemble forecast systems have useful levels of skill, that open up the possibility to develop climate services that assist agriculturalist and others in the food chain (farm suppliers, commodity traders, aid organisations) to anticipate on expected anomalous conditions. In this thesis we explore the forecast skill at various steps in the modelling chain for seasonal maize yield anomalies in East Africa. First, we analyse the skill of ECMWF System-4 (S4) climate forecasts for primary meteorological variables against gridded observations and find both potential and real skill for rainfall and temperature in typical cropping seasons in eastern Africa. However, forecast skill is a function of geographical region, season, climate variable (i.e. higher skill in temperature, rainfall, downwelling shortwave radiation in that order) and forecast lead-time, as such skill assessment should not be generalized over a large geographical area. Next we analyse correlations between reported production and anomalous weather conditions, using a range of climate indicators relevant for arable farming, such as growing and killing degree days, and rainfall amount, evenness, random independent events (unevenness), and timing during consequent maize growth phases in two case study regions. In this case significant levels of correlation and skill are revealed that open up the potential for statistical forecasting by use of climate forecast derived variables. Sensitivity of yields to climate indicators depend on geographical location, for example, higher sensitivity to rainfall is found in northern Ethiopia while in a location in equatorial-western Kenya, there is higher sensitivity to temperature indicators. At the next level of complexity we explore the use of full process based crop models forced by seasonal climate forecasts to forecast anomalous water-limited maize yield in the region, and find again potentially useful levels of skill with at least two months lead before planting, in most agricultural regions. But this again depends on regions, for example yield forecasts in Tanzania, in the season starting October did not have skill. Finally, we try to attribute skill levels to physiographic characteristics (soils, maize cultivars, geographical region etc.) and address some issues of scale of aggregation for two case study regions in Kenya and Ethiopia. The results showed that skill assessment at national boundaries and high resolution crop simulation units may inform both maize production related policy decisions at regional or national levels, and also support maize production decisions at specific cropping locations such as farm management decisions made by farmers. We conclude with a synthesis discussing further on found skill levels in relation to potential climate services aimed at the agricultural sector in East Africa and beyond.

Contents

Chapter 1

Introduction

1.1 Background

In its many forms, agriculture remains highly sensitive to climate on a range of scales e.g. climate extremes, climate variability and climate change. Agriculture is the major land use across the globe and it is of high economic, social, and cultural importance. Its sensitivity to climate is greatest in less developed countries that possess less resilience to climate impacts. Of the many climate variables, rainfall and its variability in both time and space is of utmost importance in the greater horn of Africa (GHA), an area prone to recurrent droughts and floods that have resulted into successive disasters. The success or failure of a crop growing season affects both the economic growth of East African countries as well as the livelihoods of the general populace. Impacts are experienced, whether extremes occur in the region or elsewhere largely because of relatively low food production and hence effects on availability and affordability.

Rainfall in this region is the primary water source for agriculture there being less developed structures to enable irrigation. Ground water supply for irrigation is either not fully exploited or is limited, stream flow and reservoir exploitation are limited (Auffhammer, 2011). This means that larger proportion of agricultural production is rain fed and will remain so in the near future. To meet the increasing food demands in the region, food production must be increased under these conditions yet climate variability has been found to contribute to the observed difference between actual yields and what can potentially be produced (yield gap) in Rodriguez et al. (2018). Effects of climate variability is either directly through impacts on crop physiology or indirectly via inappropriate farm management decisions. This underscores the potential to improve yields by knowing the forthcoming climate, distribution (especially rainfall and temperature) plus/or impacts on crop performance with suitable lead-time before a crop season. Further, such should enhance the potential of early warning systems. Besides enabling farmers, such should enhance the potential of relief organizations, governments, and other stakeholders to plan and adjust their activities in accordance with the expected seasonal climate.

Crop failures due to bad weather do not only affect those buying and selling in the global marketplace but also have perhaps even greater direct impact on subsistence farmers. Understanding how such extreme weather events that are also predicted to become more frequent under climate change (Boogaard et al., 2013) and its effects on both yield and total production of the world's staple food crops is thus an issue of scientific and societal concern (Auffhammer, 2011). Although understanding of seasonal rainfall anomalies is of great interest, it is the frequency of wet and dry spells within a given rainy season that is generally more important to users such as the agriculture, health, water resources, power generation and industry sectors. Thus, assessing the mean rainfall forecast over a season is not enough in some applications such as in assessing agricultural production. For example, a season may be wet overall in terms of the total amount of rainfall received but if the rain is not reasonably distributed through the season, then rainfed agriculture does not benefit (Owiti and Zhu, 2012). Good seasonal climate forecasts may help manage agriculture related risks.

1.2 Climate Forecasting and Agriculture

Forecasting seasonal climate has been possible because of the interactions of the atmosphere and slowly varying components of the climate system such as land surface, ocean surface temperatures (Hansen et al., 2011; Stockdale et al., 2010) and their teleconnections. Managing seasonal climate impacts on agricultural production in many regions of the world have been based on the prediction of El-Niño Southern Oscillation (ENSO) and its impacts on regional climates(Baigorria et al., 2008; Hoell and Funk, 2014) in regions and seasons where ENSO has influence. For example, rainfall variability over southern and eastern Africa show strong correlation to ENSO (Black et al., 2003; Goddard et al., 2001; Indeje et al., 2000; Liebmann et al., 2014; Smith and Semazzi, 2014; Omondi et al., 2013). The Indian Ocean Dipole (IOD) also drives climate variability in East Africa, either acting alone or together with ENSO variability (Black, 2005; Owiti et al., 2008; Owiti and Ogallo, 2007). Regional and local circulation patterns such as the Somalia Jet stream, monsoonal winds, topography and large inland water bodies modulate the influence of ocean processes on East African climate thus complicating seasonal forecasting in the GHA. These forcings result from internal variability of the climate system but as highlighted in Stockdale et al. (2010), other external factors such as the sunspot cycle, volcanic eruptions and greenhouse gases influence the climate to some extent. All these factors provide a basis for seasonal climate forecasting.

In the Greater Horn of Africa, operational seasonal climate forecasts are issued through the Greater Horn of Africa Climate Outlook Forum (GHACOF) convened by the IGAD Climate Prediction and Applications Centre (ICPAC); with support from the World Meteorological Organization (WMO), Global Forecast Producing Centres and other international organizations such as the International Research Institute for Climate and Society-University of Columbia (IRI), the UK-Met Office and Meteo-France (Hansen et al., 2011; Martinez et al., 2010). The forecasts are issued over a whole season and for large areas as the probability of excedence of a certain threshold, or the forecast parameter being within a range of thresholds, mostly relative to the long-term mean seasonal climate. By overlaying climate forecasts with crop production areas (see example in Figure 1.1), GHACOF also issues a statement on outlooks for agriculture for large areas generally based on expert judgment and opinions on expected impacts.

Even though these prediction methods above have been used with a relative degree of success, seasonal climate predictability is more complicated than the annual cycle of ocean sea surface temperatures (SSTs) because it is just one of the many influences of GHA's climate variability (Hansen et al., 2011), impacting predictability in some areas and seasons. For example, ENSO has influence on OND and JJAS seasons as opposed to MAM which is a major season in southern Ethiopia, in Kenya and large areas of Tanzania. Again, the atmosphere is not completely constrained and may bring rain even when oceanic conditions are not favourable (Hansen et al., 2011; Lyon and Mason, 2009). Using dynamical climate models may offer advantages in seasonal forecasting over the statistical models. Comparing GHACOF forecast skill to the UK Met Office GloSea5 dynamical model, Walker et al. (2019) found higher skill in GloSea precipitation forecasts compared to that from GHACOF; dynamical forecasts have also been found to outperform operational statistical forecasts over large areas in Australia (Rodriguez et al., 2018; Charles et al., 2015). This shows the forecast potential in dynamic models such as System-4 from the European Centre for Medium Range Weather Forecast (ECMWF) and the need to evaluate their forecast skill as a basis for application in impacts modelling.

Forecasts given over large areas may not capture the intricate interactions between large-scale simulations and local features that influence climate and hence their use to assess impacts on agriculture in smaller areas is limited. Forecasts given in terms of mean seasonal rainfall or temperature may not favour explicit use in applications such as crop production since the distribution of weather and timing over a crop growing season impacts yields more compared to the total or average seasonal weather. Crop growth over a growing season is determined by interactions of the various weather elements such as moisture availability and daily temperatures on a daily time scale. Thus,current operational forecasts cannot be used to explicitly assess crop growth but may give an estimation of expected yields. This has been useful. For example, a stronger correlation between maize yields and the Pacific sea surface temperatures (SSTs) associated with ENSO than to seasonal total rainfall has been found in Zimbabwe (Cane et al., 1994). Climate change further complicates the use of statistical models to either forecast oncoming season or to predict crop yields because the world is slowly moving into climate regimes not experienced historically in terms of atmospheric temperatures, climate variability, and the frequency of extremes. Globally, a non uniform relationship has been observed between maize yields and ENSO signal in the current climate. Yield impacts are more in ENSO-teleconnected lower income countries due to their lack of resilience to ENSO signals (Ubilava and Abdolrahimi, 2019). The ability to predict seasonal climate before the start of a planting season raises prospects of their use in estimating the coming season's crop yield by use of dynamic crop models (Shin et al., 2009). In addition, reliable seasonal forecasts with suitable lead-time before planting would enable adoption of farm management practices as dictated by the oncoming growing season climate.

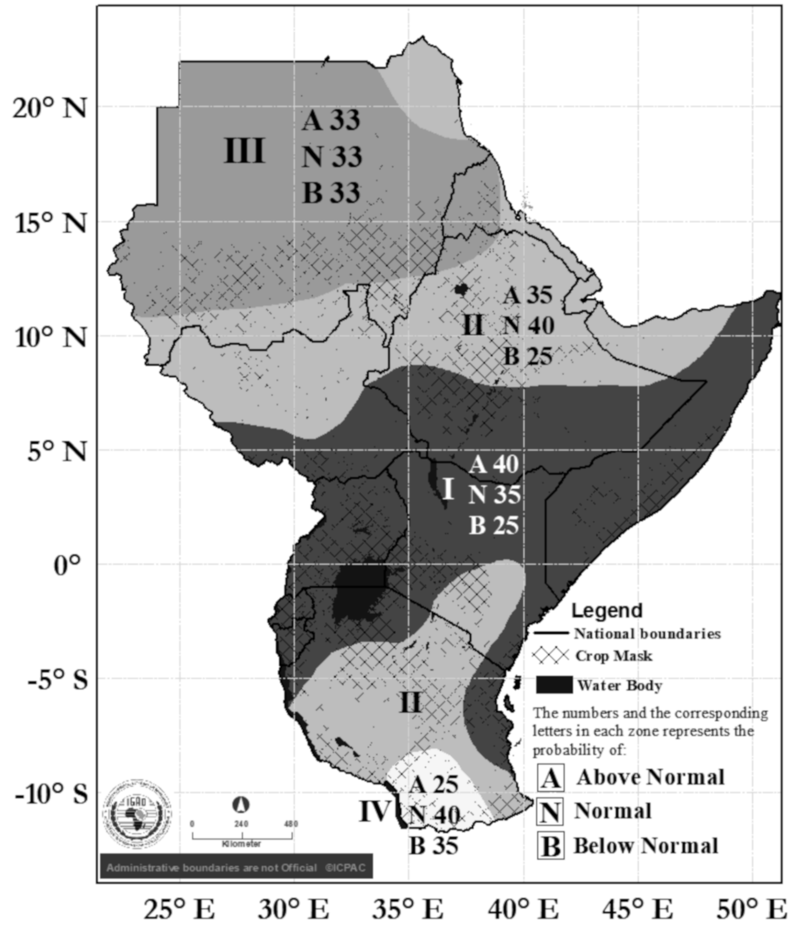


Figure 1.1: Map showing GHA October-December 2018 probabilistic rainfall forecast overlay with cropping areas to show crop performance outlook for agriculture and food security sector (source: ICPAC (2018))

1.3 Linking Dynamical Seasonal Climate Forecast and Crop Models

An alternative to maize production assessment is to use process based dynamic seasonal forecasts provided by GCMs and crop simulation models. Direct application of GCM forecasts in agricultural impact modelling are limited (Hansen et al., 2006) requiring first a downscale to suitable resolutions by either regional climate models (RCMs), weather generators (Cantelaube and Terres, 2005; Semenov and Doblas-Reyes, 2007) or either by a range of statistical downscaling or bias correction techniques. Because of the uncertainty of climate forecasts over a season, multi-model ensemble (MME) forecasting approaches are employed to generate more reliable probability forecasts of seasonal climates rather than the use of a single model (Doblas-Reyes et al., 2009; Palmer et al., 2004) a phenomenon that may hold for agricultural models as well (McIntosh et al., 2005; Martre et al., 2015; Asseng et al., 2013).

Seasonal climate forecasts have been used in agricultural impacts modelling globally with varied results suggesting variability due to factors like spatio-temporal scales, surface heterogeneity, crop management practices, and model initialization amongst others (Jones et al., 2000; Shin et al., 2009). An important factor perhaps is climate predictability, assuming that skilful seasonal forecasts should bear similarly skilful agricultural yield predictions. But driving agricultural impact models with skilful seasonal climate forecasts may not guarantee good yield forecasts (Baigorria et al., 2007; Semenov and Doblas-Reyes, 2007; Shin et al., 2010), the reverse i.e. better skill in the crop forecast than in the meteorological forecast has also been reported (McIntosh et al., 2005). In addition, the time of the year, and lead-time in which a forecast is useful varies depending on the crop and region (McIntosh et al., 2007) that in turn depends on cropping calendars.

It would be useful therefore to know the forecast skill of both climate forecasts and maize yield forecasts as a function of geographical area, season, and forecast lead-time. The findings of this research will improve modelling by identifying suitable lead times and best methods to drive the impact models besides detailing the crop characteristics over the study area. Better so, the lead-time of useful forecasts to enable change in farm management decisions. For successful crop yield forecasts in a locality, it becomes necessary to know the uncertainties involved to facilitate apportionment of confidence levels to each model. Reducing uncertainties should improve prediction skill and the use of ensemble forecasting techniques helps provide information on forecast uncertainties. Further, the ability to simulate impacts of extremes improves the capability to deal with such events presuming a pro-active response.

Reliable agricultural impacts predictions are useful for seasonal early warning and adaptation policy advisories. Seasonal probabilistic yield and production forecasts either in advance of harvest or sowing is important if used appropriately, with understanding of their capabilities and limitations. This is not enough for the farming communities whose ability to cope with the constraints and opportunities of climate variability need enhancement (Hansen and Ines, 2005; Skees et al., 1999). A climate information service should then deliver understandable impact predictions to stakeholders, considering locations in which particular methods or models are applicable. Findings of this research should augment the information provided by GHACOF by providing explicit quantitative impacts on agriculture by use of dynamical climate forecasts and crop simulation models.

1.4 Objectives of the Study

The broad objective of this project is to test whether seasonal climate forecasts in combination with crop model can be a useful addition to the existing agricultural extension services or may form a basis for establishment of agricultural climate service. This assessed by making use of currently available seasonal climate forecasts form one of the global forecast producers, and a physical process based crop model. It assesses skill in the climate-yield model chain, assess growing season climate variable characteristics as a basis for statistical modeling and the impacts of model grid resolution, level of aggregation, soils and geographical region in maize yield simulation. The specific research objectives were:

1. Assess the skill of GCM and bias corrected seasonal climate forecasts over East Africa via comparative analysis of hindcasts with observational data for the period 1981 to 2010

2. Assess maize yield predictive skill of a crop simulation model through both baseline and hindcasts validation for the period 1981 to 2010

3. Asses the influence of choice of cultivars, and spatial aggregation on predictability of seasonal forecasts of maize production

4. Assess the important growing season climate characteristics or indices that influence maize yield predictions, and their predictability

The above objectives are addressed by use of GCM ensemble seasonal climate forecasts from one of the global forecast producing centres, gridded pseudo historical climate observations, historical official national yield observations, and a field-scale crop simulation model configured and upscaled for regional crop simulation.

1.5 Methodological Approach

To answer the above questions, long time series of daily historical observed climate data for 1981-2010 from Watch Forcing Data ERA-Interim (WFDEI) (Weedon et al., 2011) will be used as a reference observation. Hindcast ensemble (15-ensembles) seasonal forecasts provided by European Centre for Medium Range Weather Forecasts (ECMWF) (Molteni et al., 2011), is post processed, bias corrected and verified against historical observations by use of a range of verification measures already developed and used in meteorological communities. Even though forecast skill have already been computed by the forecasting centres at their native grid box level, in this research, an in-depth regional analysis for relevant seasons and forecast lead times is performed. Forecast skill of variables that are important for crop growth i.e. rainfall, temperature, and incoming shortwave radiation are assessed for relevant cropping seasons in East Africa.

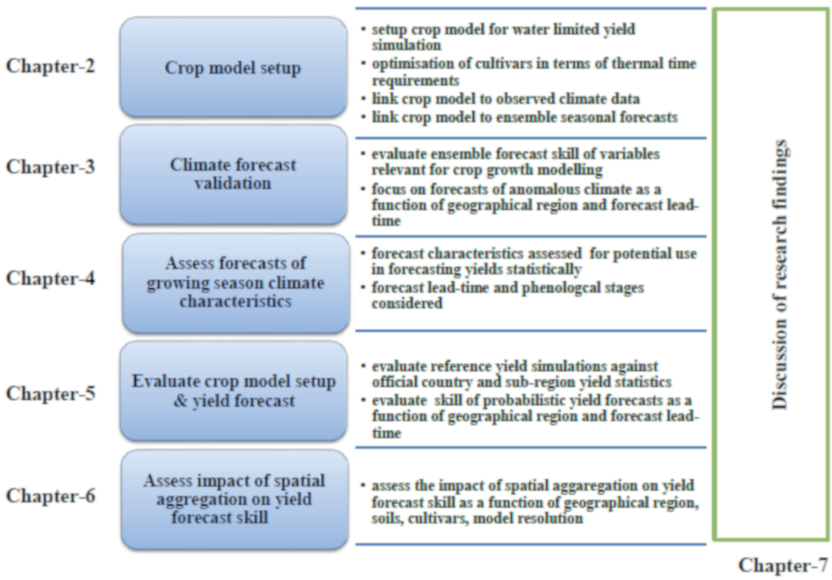


Figure 1.2: Research steps

Hindcasts of crop yield are created by driving both the World Food Study (WOFOST) crop growth model (Boogaard et al., 2013; Van Diepen et al., 1989; Keulen and Wolf, 1986; Van Diepen et al., 1988; Supit et al., 2010) with the reference gridded pseudo observations (WFDEI) and compared to official maize yield statistics from Food and agriculture organization (FAO), and data from each of the country's agricultural statistics authorities. This enables assessment of the goodness or correctness of gridded reference yields relative to the official observed yield statistics. Maize production forecasts are generated by driving the WOFOST with each of the 15-climate forecast ensemble members, resulting in fifteen yield ensemble forecasts thus enabling sampling of forecast uncertainties. WOFOST is a field scale crop simulation model but setup and configured in this study for regional crop simulation based on the FAO landuse map and with a resolution similar to its landuse cells ($\approx 0.1°$ grid) with region specific cultivars. Only water limited yield maize production is simulated.

Comparisons and assessment of yield prediction skill is at sub-national levels for chosen agricultural regions subject to availability of at least ten years observed data. It must be noted here that a model, being a simplification of reality is not expected to simulate results that completely agree with the actual observed values, but focus is on yield anomalies. In any case, the 'actual' observed yields from the national statistical bureaus of the individual countries are themselves estimations, not completely accurate, and the methods to compile the statistics vary among countries. The essence of comparison and improvement will be to provide adequacy of the relative, year-to-year variations in simulated yields, rather than of absolute production; to provide adequacy of forecasted yield percentiles (e.g. terciles or quintiles). A set of established probabilistic performance statistical measures (Murphy and Wilks, 1998; Willmott and Matsuura, 2005) are used to verify both climate and yield forecasts. The same probabilistic forecast skill measures from the weather and climate forecasting communities are applied to maize yield predictions. Validation of yield forecast is at both the crop model grid and sub-regions of the study area by use of case studies regions.

Seasonal climate forecasts possess uncertainties because of the long forecast duration. It is also known that the total or average season climate (e.g. rainfall) may not influence variability in yields as much as the distribution over the growing season would. Growing season climate characteristics that influence yield variability are identified and their predictability assessed as a function of forecast lead-time. In summary, this research sought to identify an optimal lead-time, geographical region and period skill that would form the basis for an integrated forecasting system from climate to maize production.A summary of the research approach is shown in Figure 1.2.

1.6 Overview: Agriculture and climate in the study area

1.6.1 Ethiopia

Ethiopia has a wide range of ecological diversity ranging from tropical to temperate climates with a land area of approximately 1.2 million Km^2 of which, ≈61% is arable. Ethiopia's topography ranges in altitude from −125 masl in the north East to 4620 masl in the North-West. Average temperatures range from 15°C over the mountains to 35°C in the lowlands (Jury and Funk, 2013). Major cropping areas are in the sub-humid, humid and moist-semi arid zones while major cereals are grown in regions with elevation of 1800 to 3000 masl, average annual rainfall of 950mm to 1500mm and mean annual temperature of 11 to 21°C. Maize particularly is grown in between 1500 masl to 2200 masl in almost all regions of the country but with variations in production due to moisture stress, low soil fertility, lack of improved cultivars, pests, diseases, weeds, among others (Fantaye, 2020). Soils of Ethiopia comprise 23% Nitosols, 19% Cambisols and 18% Vertisols in more than half of the arable area in different ecological zones. Agriculture is a major economic activity in Ethiopia contributing to ≈43%

of the national Gross Domestic Product (GDP) and employs 85% of the working population. Small holder farming under rain fed conditions predominate agricultural production systems, and >95% of national output is generated by subsistence farmers owning less than 1 ha of cultivated land with poor soil fertility(FAO, 2020a; Taffesse et al., 2012; Kassie et al., 2014; Kassie, 2014; Komarek et al., 2019). Agriculture in Ethiopia highly depend on rainfall and its characteristics such as onset, duration, amount and distribution determines performance of the sector and as a consequence, the national economy.

1.6.2 Kenya

Kenya has a total land area of ≈576076 Km^2, out of which about 17% has high to medium agricultural potential for intensive crop cultivation. The rest is classified as arid and semi-arid (ASAL) lands. More than 7 million people derive their livelihoods from the ASALs while the remaining population live in the high agricultural potential areas or in cities. Rainfall amounts and temporal distribution determine cropping systems because agriculture is mainly rain fed often resulting in serious impacts due to recurrent cycles of floods and droughts. Less than 7% of cropped land in Kenya is under irrigation (Boulanger et al., 2018; Adimo, 2020). Kenya is characterized by two distinct rainfall seasons (i.e. the long rains of March-May and short rains of October-December) though minor season from June-August is also seen in the western and coastal regions resulting in regions with single-crop and double-cropping systems. Soils vary from sandy-to-clay, shallow-to-very deep, low-to-high fertility due to varying geology, relief and climate. Major soils used for agriculture are ferralsols, vertisols, acricols, lixisols, luvisols, and nitisols (Omuto, 2013; Adimo, 2020). Agriculture is a major economic activity in Kenya providing ≈ 33% of the national GDP directly as of 2016 (GOK, 2019), another 27% indirectly through linkages with other sectors, and employs 40% of the total population and supports ≈ 80% of the rural population (Adimo, 2020; FAO, 2020b). It accounts for 60% of export earnings. This sector has been identified in Vision (2007) as a key sector expected to drive Kenya's economy to an annual growth of approximately 10% by 2030, and to deliver to other regional and global commitments such as the African Union Agenda 2063, Sustainable Development Goals (SDGs) and the Comprehensive Africa Agricultural Development Programme (Boulanger et al., 2018). Agricultural transformations are needed to achieve these, especially by improving smallholder farmer's productivity. Farming systems range from small-scale to large-scale mechanized enterprises. Land ownership in the high potential areas range from 0.5 to 10 ha. Today, there are ≈4.5 million small scale farmers in Kenya accounting fro 63% of the national agricultural output on approximately 90% of land under agriculture. Out of this, 3.5 million are crop farmers (GOK, 2019).

1.6.3 Tanzania

Tanzania has a tropical climate and is divided into four climatic zones i.e. the hot humid coastal plains, semi-arid central plateau, high rainfall lake region, and the temperate highlands. Hottest periods are in November to February with a temperature of 10°C to 31°C, while the cold period is from May to August with temperatures of 15˚C to 20˚C (Makio, 2020). The annual mean temperature range from 17˚C to 27˚C depending on location. Relative humidity range between 50% and 80% in the whole country with the main moisture sources being the Indian Ocean, Lake Tanganyika (to the west) and Lake Victoria to the north (Cioffi et al., 2016; Makio, 2020). Tanzania experiences two rainfall regimes, uni-modal (musuni) from October to April in southern, central and north western parts of the country; and a bi-normal pattern of long rains(Masika) from March to May (MAM) and the short rains (vuli) from October to December (OND) in the lake region and extending east to the coast. The bi-modal pattern is caused by seasonal migration of the ITCZ. MAM rains are associated with northward, and OND rains with the southward migration ITCZ (Borhara et al., 2020; Cioffi et al., 2016). Rainfall cycles result from annual cycle of monsoonal winds in combination with Indian Ocean sea surface temperature (SST) and modification of large-scale atmospheric flows by local topography. Tanzania has a range of soil types highlighted in (Makio, 2020), but not all types are suitable for agricultural activities. High potential agricultural regions of Arusha, Kilimanjaro and southwest highland regions are dominated by volcanic soils. Light sandy soils along the coast are suitable for grazing in the rainy seasons; poor soils of granite origin in the mid-west; red soils in the central plateau is used for forage in both the dry and rainy seasons; and vertisols soils that is widespread and used for forage in the dry seasons. FAO-UNESCO Soil Classification System however classifies Tanzania soils into nineteen different types discussed in Mlingano (2006) There also are ironstone soils mainly in Kagera, Kigoma, Sumbawanga are poor and acidic but their productivity can be improved by use of farm management practices such as nutrient application, and mulching.

Agriculture is a major factor in the economy of Tanzania contributing to about 24% of GDP, accounts for 24% of exports and employ almost 75% of the working population. Out of a total land area of 94.5 million hectares, 44 million hectares are classified as arable but only 26 millon hectares is under use. Area under irrigation by 2013 is less that 5% of the cultivated land. About 23% of the arable land is under cultivation dominated by food crop production (World-Bank, World-Bank; Makio, 2020). Agriculture is mainly rain fed, and this has affected agricultural production due to increasingly unreliable and irregular weather conditions.

1.7 Summary; study area

The economies of these three countries and millions of its populace are dependent on agriculture with a majority of the rural poor involved in rain fed agricultural production. Because many poor households are involved in agricultural production, there exist a potential for poverty alleviation with proper interventions since the sector growth is strongly tied to climate variability. Knowledge of expected weather conditions in advance of a cropping season may help mitigate against weather vagaries by influencing some farm management decisions that improve profitability of agriculture by for example, increasing gains in favorable seasons or reducing losses in dry seasons (Parton et al., 2019). Luckily, seasonal climate over this region is predictable some months in advance due to the interactions of the atmosphere with slowly varying ocean and land components of the climate system (Thomson et al., 2006; Doblas-Reyes et al., 2013). A major source of predictability that forces climate anomalies globally is the El-Niño Southern Oscillation (ENSO), it is predictable with lead-times of up to 12-months (Harrison, 2005; Jan van Oldenborgh et al., 2005; Charles et al., 2012; Doblas-Reyes et al., 2013). It is known to have strong correlations with East African rainfall (Goddard et al., 2001; Indeje et al., 2000; Semazzi et al., 1988). The Inter-Tropical Convergence Zone (ITCZ), and the Indian Ocean Dipole (IOD) among other features also influence East African seasonal climate (Goddard et al., 2001; Owiti and Ogallo, 2007). The possibility to forecast seasonal climate with useful skill in his region holds the potential for use by a range of decision-makers to better manage climate risks. Predictability of climate forecasts also provide an opportunity for new technologies to also deliver sectoral impacts information. Potential skill in predictability of East Africas'climate makes the region suitable for exploring the seasonal climate-agricultural impacts modelling axis with assumption that skillful climate forecast should likewise result in skillful impacts prediction at lead-times that would influence decisions.

1.8 Thesis Outline

This report consists of seven chapters; chapter 1 provides the context of this research; highlighting the objectives plus research approach. Chapter 2 deals with the setup and configuration the crop growth model. The chapter is important because the model is designed for field scale use but in this study, it is configured for a regional scale simulation using grids of 0.1° by 0.1° degree as the basic simulation unit. The following four chapters will address each of the research objectives.

Chapter 3 identifies from the seasonal climate forecast systems, regions, seasons ((i.e. MAM, JJAS, OND, JF seasons) and lead-times of skillful climate forecasts. It will indicate whether there is a necessity to bias correct the climate forecasts before using them to drive the crop. The chapter will identify optimum climate forecast lead-time for optimum skill and reliability before use in a crop model. It focuses on climate variables relevant for crop model and yield simulation.

It is known that average seasonal climate conditions plus distribution of weather during a crops' growth season influence yield variability. Chapter 4 identifies the important weather characteristics or indicators in the growing season and consequent crop growth stages that explain yield variability. Prediction skill of identified indicators could provide an alternative yield forecast system using statistical methods rather than the use of full forecast ensembles.

Chapter 5 highlights the results of crop model set up, yield simulations and yield forecast skill. It gives a comparison of the reference simulated yields to the official observed yield statistics, and model simulated cropping dates to the observed The chapter also indicates cropping seasons, geographic regions, and lead times in which maize yields can be reliably predicted.

Chapter 6 assesses the influence of yield aggregation on forecast skill. This is done based on horizontal grid resolution, physiographic characteristics (i.e. soils, maize cultivars, geographical region etc.) and address some issues of scale of aggregation for two case study regions in Kenya and Ethiopia.

Chapter 7 provides a synthesis of the findings of this research in a wider context. It elaborates on the methodological strengths and limitations to the research setup. It discusses the findings from all chapters and implications to the scientific community and society. It discusses further on skill levels in relation to possible agricultural sector climate services in East Africa. It gives recommendations on further research, and how this methodological setup may fit in existing early warning systems and possible use in climate risk management.

Chapter 2

Implementation of WOFOST for Yield Prediction over East Africa

2.1 Introduction

Sub-Saharan Africa (SSA) is one region of the world that is and will be highly impacted by climate change. Agricultural sector is already affected due to over reliance on weather for agriculture i.e. Africa is already experiencing changes in average temperature, rainfall amounts and patterns, frequency and intensity of extreme events. This has or will consequently impact agricultural systems for example, length of growing seasons, yields (Kotir, 2011) and possibly the traditional crop calendars. These changes in climate also affect east African region. With an increasing population and reducing per capita food production, the region needs to boost food production in order to feed the populace (van Loon et al., 2018). A number of methods such as expansion of cultivated areas, intensify irrigation systems, improved soil fertility, improved crop varieties, or optimum use of climate information may improve yields.

Infertile soils has been identified to hinder maize yields by the highest political leadership in the continent thus a resolve by Africa Fertilizer Summit to increase fertilizer consumption from 8 $\mathrm{Kg\,ha^{-1}}$ in 2006 to 50 $\mathrm{Kg\,ha^{-1}}$ by 2050 (AGRA, 2019; Declaration, 2006). One of the reasons identified in Bonilla Cedrez et al. (2020) for low fertilizer uptake in SSA is that "variability of rainfall makes it too risky to invest in fertilizers and insurance programs may be needed to support fertilizer use". Agriculture in this region is largely rain fed and sensitivity to weather aggravate farmer's challenges (Tesfaye et al., 2017). In addressing yield gap in SSA, Hillocks (2014) notes that yield gap (i.e. the difference between potential and actual yields) would be difficult to close under rainfed conditions. Optimal use of of climate information may help farmers to cope.

To explore the yield-climate interactions, or how other factors like fertilizer application affect crop models using a range of approaches can be used. In weather-crop simulation, available approaches are based on crop physiology and climate understanding (i.e. climatological, water-stress models, dynamic crop-weather models); and based on type and purpose (i.e. statistical, mechanistic, deterministic, dynamic, static,simulation, descriptive and explanatory models). Crop models have been used globally as a tool to explore effectiveness of differing crop management practices, to study effects of inputs, explore climate change adaptation options to minimize climate risks,and to study crop growth processes. Crop model use in the context of climate change impacts on food security has been strongly simulated in recent decades amongst others through the Agricultural Model Intercomparison and Improvement (AgMIP) Programme (Rosenzweig et al., 2013). This research explores the use of seasonal climate forecasts and a crop model to forecast maize yields. A crop model, World Food Studies crop simulation model (WOFOST) is used. The following sections will explain the setup of such model and experiment.

2.2 Methods and data

2.2.1 Model Description

We use the World Food Studies crop simulation model (WOFOST) version in the Python Crop Simulation Environment (PCSE/WOFOST) implementation. The PCSE is a Python package for building crop simulation models. It provides the environment to implement crop simulation models, tools for reading ancillary data (weather, soil, agromanagement), and the components for simulating biophysical processes such as phenology, respiration and evapotranspiration (de Wit et al., 2019). Also included in PCSE is the LINTUL3 crop model.The model is provided with full source code and documentation through public repositories (de Wit et al., 2019) and details are available at https://pcse.readthedocs.io/en/stable/index.html.

WOFOST is a simulation model for quantitative analysis of growth and production of annual crops. It is a mechanistic crop growth model that describes plant growth by using light interception and CO_2 assimilation as growth driving processes and by using crop phenological development as a growth controlling process. WOFOST was extended with the nutrients routines similar to the ones in in the Lintul model. These routines are based on Shibu et al. (2010). It can simulate potential yield situation, water limited, and a combined water and nutrient limited situation. WOFOST is photosynthetically driven and simulates the growth and production of annual crops using a range of daily physiological processes in response to weather, soil types, soil moisture conditions,as defined by crop cultivar characteristics. Physical processes include light interception, photosynthesis, respiration, evapotranspiration, assimilate partitioning, leaf area dynamics, phenological development, and root growth. Not implemented are the effects of pests, weeds, farm management practices de Wit et al. (2010); Boogaard et al. (2013). The model can simulate either potential yields or water-limited yield. Potential yield simulation is where crop growth is driven purely by temperature, day-length,solar radiation and cultivar characteristics assuming no water nor other growth limiting factors. In water-limited yield simulation, crop growth is limited by water availability but assumes adequate nutrient supply de Wit et al. (2010) and hence influenced by rainfall, soil type and field topography but assumes adequate nutrient supply (Boogaard et al., 2013). WOFOST was originally developed to simulate crop yield for a single location with homogeneous cultivars, soil characteristics and weather (Boogaard et al., 2013; Van Diepen et al., 1989; Supit et al., 2010) but in this study, it is set up for a regional simulation.

2.2.2 Input Data

2.2.2.1 Crop data

Many approaches have been employed to estimate crop sowing dates in regional yield simulation studies. This includes the use of observed panting dates and expert advice, use of rainfall and characteristics at the start of planting season, use of cumulative rainfall and soil moisture, and optimization of sowing dates on maximum yields plus minimum variability in yields. In this study, we use a method that combines the "optimization of sowing dates on maximum yields" and "optimization on climate and crop specific characteristics". This procedure has been used and de-tailed in Wolf et al. (2015). We start with crop calendars from FAO (http://www.fao.org/agriculture/seed/cropcalendar/welcome.do) and Sacks' calendar described in Sacks et al. (2010)), and standard tropical maize varieties for crop modelling compiled in Van Heemst (1988) and Van Diepen et al. (1988).

2.2.2.2 Land use and soil data

Soil input data are derived from the International Soil Reference and Information Centre-World Soil Information (ISRIC-WISE) database (Batjes, 2012). The database includes information on soil physical characteristics, root depth, and landscape characteristics such as elevation, slope gradients, and slope aspects. Soil properties such as wilting point and field capacity are estimated with the pedotransfer functions from Saxton et al. (1986).Maize growing areas are determined based on FAO land use maps (Fischer et al., 2008).

2.2.2.3 Historic weather data

Since we want to use WOFOST over a large spatial domain where weather stations are relatively sparse and often produce incomplete records, we revert to gridded weather datafrom other sources that do have the required continuity in the space and time domains. The Water and Global Change (WATCH) project (Boucher and Best, 2010; Harding et al., 2011; Weedon et al., 2011) prepared a meteorological forcing data, the

WATCH Forcing Data (WFD) making use of ERA-40 reanalysis (Uppala et al., 2005) from the European Center of Medium range Weather Forecasts (ECMWF) for the period 1958-2001. WFD was based on interpolation of ERA-40 reanalysis to 0.5°× 0.5° resolution followed by elevation correction of surface meteorological variables and monthly bias correction using gridded global land observations at 0.5-degree resolution. The WATCH Forcing Data methodology applied to ERA-Interim (WFDEI) was processed using a similar methodology to WFD but based on ERA-Interim reanalysis. ERA-40 and ERA-Interim reanalysis details are in Weedon et al. (2014) and Dee et al. (2011).

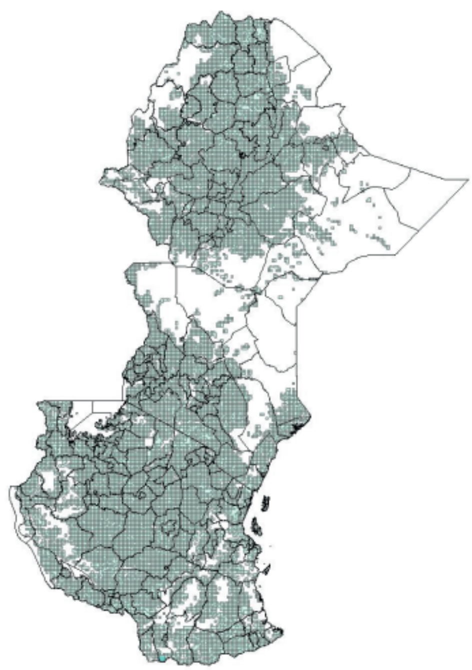


Figure 2.1: FAO land use map showing rainfed agricultural regions of the study area at 0.1°$times$0.1° resolution

Briefly, ERA-Interim compares GCM modeled state of observations at 6-hourly interval including allowance for their exact time stamp as opposed to ERA-40 reanalysis (Weedon et al., 2014) leading to steady updates that are consistent to the observed. ERA-Interim incorporates more satellite derived products, atmospheric soundings, and surface observations hence providing an improvement over the ERA-40. As elaborated in Dee et al. (2011), there is an improvement in representation especially of hydrological variables i.e. improved humidity analysis, assimilation of satellite passive microwave data for total column water vapor in areas affected by clouds and rain, and assimilation of satellite derived snow extent. ERA-Interim has a reduced Gaussian

Grid Spectral model resolution of T255 (approximately 0.7° at the equator). From WFDEI, precipitation, air temperature and downwelling shortwave radiation are used in this study. Processing applied to these variables are explained in the following paragraph but details are available in Boucher and Best (2010); Weedon et al. (2011, 2014).

This study uses the WFDEI rainfall data corrected against CRU data(Harris et al., 2014). Precipitation totals,ERA-Interim rainfall/precipitation ratio,and rainfall gauge are bias corrected using CRUTS3.1.01/TS3.21 while CRU TS3.1 is used to correct the number of wet days. In regions with large elevation differences between ERA-Interim and CRU grids, differences in precipitation phases are corrected via the procedure elaborated in Weedon et al. (2014). Briefly; for each grid box and each calendar month, over in the period 1979-2009, minimum temperature (*tmin*) during rainfall and maximum temperature (*tmax*) during snow fall are stored. Each phase temperature extreme is obtained from the library of eight 3-hourly steps $\times$ 30-days(a month) $\times$ 31-years. For each grid box and 3-hour time step, precipitation phase was switched if combination of phase with elevation and bias corrected air temperature (*tair*) lay beyond a phase temperature extreme. In most grid boxes and time steps precipitation phase remained unswitched because the spatial resolution of elevation in the ERA-Interim and CRU are fairly similar. Two-meter air temperature data in ERA-Interim were bias corrected for elevation by use of environmental lapse rate. Average monthly temperatures and diurnal temperature range were bias corrected using CRU TS3.1/3.21 data. Downwelling short wave radiation is bias corrected using CRU TS3.1/3.2 cloud cover and effects of interannual changes in atmospheric aerosol loading. Monthly aerosol correction for downwelling shortwave radiation fluxes was done separately for clear and cloudy sky. corrections for aerosol loading were implemented after correction for distances between monthly modelled and observed cloud cover. The WFDEI is used as the best possible proxie, continuous in space and time, for observed historic weather data.

2.2.2.4 Climate Forecasts

S4 re-forecasts starting 1981 to 2010 is evaluated. This is a fully coupled ocean–atmosphere GCM based on the Integrated Forecast System (IFS c36r4) atmospheric model at a resolution of TL255 (approximately 0.75° horizontal resolution) and 91 vertical levels. Forecast initialization starts on the first day of each month with fifteen perturbed initial conditions giving 15-ensemble members, each providing a forecast 7-months into the future. A combination of atmospheric singular vectors and an ensemble of ocean analysis provide a means to create the perturbations (Molteni et al., 2011).

2.3 Setup

In order to use WOFOST for yield simulation, it was necessary to establish units of simulation [Soil mapping units (SMUs)]. These are grid cells of more or less homogeneous weather, soil and crop characteristics. A number of data sets were required to establish SMUs i.e.

- WFDEI at 0.5°×0.5°horizontal grid resolution
- Climate forecasts (*S4*) at 0.75°×0.75° horizontal resolution (15-ensemble members)
- FAO land use cells at 0.1°×0.1° horizontal resolution
- Soil map from ISRIC-WISE database

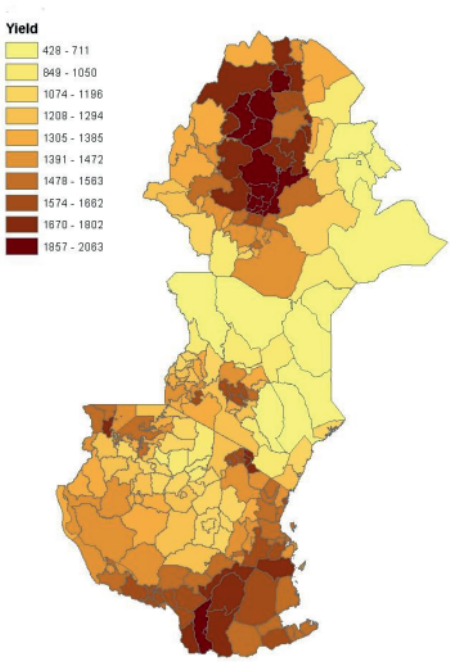


Figure 2.2: Average maximum WFDEI simulated yields (in Kg ha^{-1}) over the period 1980-2011

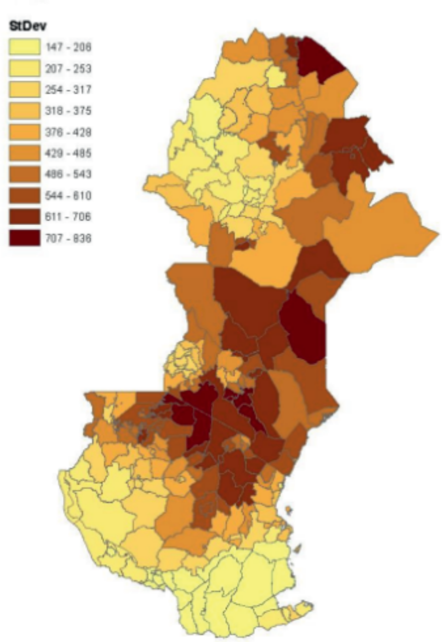


Figure 2.3: Average standard deviation of yields (in Kg ha^{-1})

From land use map, we selected only land use cells where agricultural landuse ≥5% and disregarded irrigated areas. We assumed then that the selected land used cells are rain fed. The resulting FAO land use map (shown in Figure 2.1) was overlayed with soil map and soil properties assigned to each cell using pedo-transfer rules from Saxton et al. (1986). The next step was to assign climate data to each SMU. The resulting map containing land use and soil information was thus overlayed with both WFDEI and *S4*, WFDEI having been remapped to *S4* resolution of 0.75°. This was done to preserve the GCM *S4* simulations. The forecasts were bias corrected using the quantile-mapping method A.1. For the baseline (WFDEI), we attributed to all land-use cells within a WFDEI (0.75°×0.75°) box the weather that belongs to that grid. Also assigned to each is the weather from *S4*, each grid having 15-ensemble members. Weather variables assigned to each grid include maximum temperature, minimum temperature, wind, rainfall and irradiance. Since each land use cell has required soil and weather data, the next step was to establish sowing dates and crop varieties.

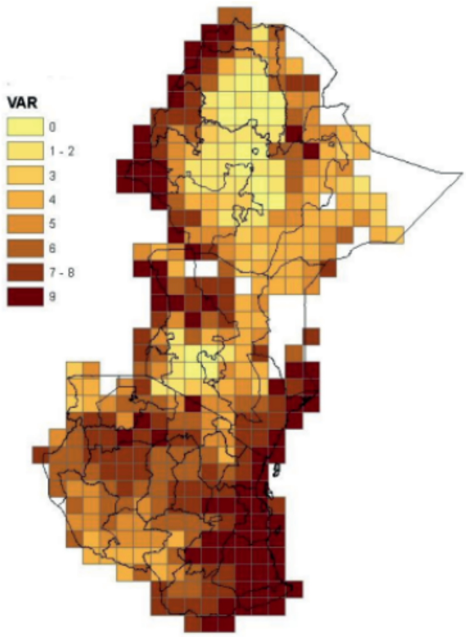


Figure 2.4: Derived nominal crop varieties based on optimized TSUMs

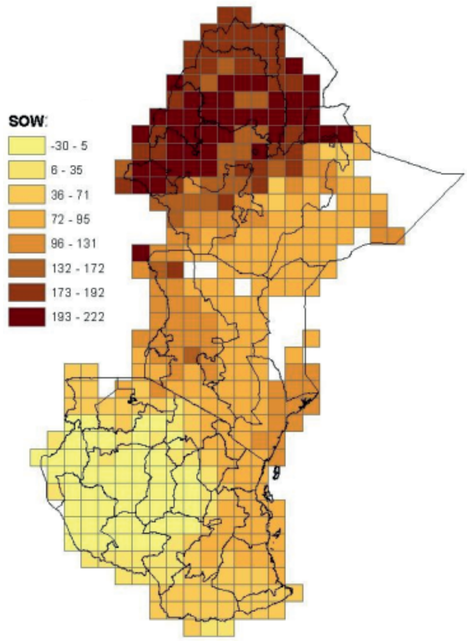


Figure 2.5: Optimal sowing dates (units day of the year)

We did not have specific crop varieties for the region, so we first chose cultivar parameters detailed in Table-6 of Van Heemst (1988) and allocated this to the simulation units. Thermal unit approach is used to simulate phenological development. Each phenological of teh crop is expressed by *DVS* Next, we allocated to each simulation unit the crop calendars in Sacks et al. (2010). This is a gridded map resulting from digitizing and georeferencing of observed dates and a derivation of climate statistics such as the average temperature at which planting occurs in each region. To each simulation unit, we allocated dates that coincide to the first seasons of the year. Assuming availability of $60\text{Kg}\,\text{ha}^{-1}$ of Nitrogen (N) during the growing season. This assumption gives simulated yields same range as the statistical yields.

Next, we derive optimal sowing dates and crop varieties for each region using the WFDEI climate data for 1980 to 2011. Yields are simulated for maize varieties planted at every grid point every 10 days, starting 90 days before and end 90 days after the calendars in Sacks et al. (2010) using 10-11 TSUMS. TSUMs refer to thermal heat units for each crop variety required for maize to advance in phenology from emergence to anthesis (TSUM1) and from anthesis to maturity (TSUM2). It is obtained by cumulative summing of the difference between daily temperature and the base temperature (10°C)for each day that daily temperature is above the base temperature. Thermal time accumulation in WOFOST is based on planting dates and a base temperature on 10°C and upper limit of 30°C. We then selected the sowing date-TSUM-development stage combination that resulted in maximum average yields (Figure 2.2) and lowest deviation (Figure 2.3) over the period 1980-2011. The resulting crop parameters for the varieties (Figure 2.4) and sowing dates (Figure 2.5) are then attributed to each simulation unit. TSUMS corresponding to these maize varieties are in Table 5.1. The dates and varieties are then fixed for both historical and forecast yield simulation. harvest date are not fixed but are determined by crop-soil-weather characteristics during the growing season.

Simulated yields are stored in a database at the resolution of SMUs (0.1°×0.1°) but these can be aggregated to any other resolution (e.g. 0.5°), national or sub-national administrative boundaries by weighting against the fraction of cultivated area. The NUTS concept used and described in Supit and Van der Goot (1999) and Boogaard et al. (2013) is adopted in this study. NUT0 refers to national administration units while NUT1 refers to sub-national region in which also official yield statistics are collected. A validation of NUT0 level model simulations are detailed in chapter **??**. While a conceptual diagram of the experiment setup is shown in Figure 2.6. Simulations obtained from the approach described in this chapter form the basis of Chapters 4 to Chapter 6.

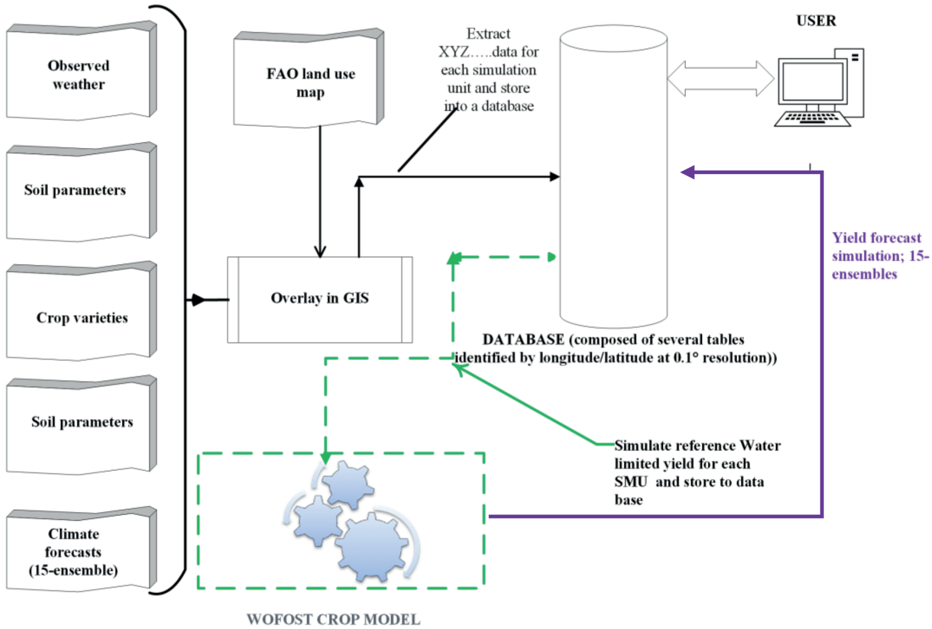


Figure 2.6: A conceptual diagram of experiment design

Chapter 3

Skill of ECMWF System-4 Ensemble Seasonal Climate Forecasts for East Africa

Abstract

This study evaluates the potential use of the ECMWF System-4 seasonal forecasts (*S4*) for impact analysis over East Africa. For use, these forecasts should have skill and small biases. We used the 15-member ensemble of 7-month forecasts initiated every month, and tested forecast skill of precipitation (*tp*), near-surface air temperature (*tas*) and surface downwelling shortwave radiation (*rsds*). We validated the 30-year (1981–2010) hindcast version of *S4* against the WFDEI reanalysis (WATCH Forcing Data ERA-Interim) and to independent relevant observational data sets. Probabilistic skill is assessed using anomaly correlation, ranked probability skill score (RPSS) and the relative operating curve skill score (ROCSS) at both grid cell and over six distinct homogeneous rainfall regions for the three growing seasons of East Africa (i.e. MAM,JJA and OND). *S4* exhibits a wet bias in OND, a dry bias in MAM and a mix of both in JJA. Temperature biases are similar in all seasons,constant with lead-time and correlate with elevation. Biases in *rsds* correlate with cloud/rain patterns. Bias correction clears biases but does not affect probabilistic skills. Predictability of the three variables varies with season, location and lead-time. The choice of validating dataset plays little role in the regional patterns and magnitudes of probabilistic skill scores. The OND *tp* forecasts show skill over a larger area up to 3 months lead-time compared to MAM and JJA. Upper- and lower-tercile *tp* forecasts are 20–80% better than climatology. Temperature forecasts are skillful for at least 3 months lead-time

This chapter has been published as:
Ogutu G.E., W.H. Franssen,I. Supit,P. Omondi and R.W. Hutjes,2017: Skill of ECMWF system-4 ensemble seasonal climate forecasts for East Africa. *International Journal of Climatology*, **37** (5), 130-146.

and they are 40–100% better than climatology. The *rsds* is less skillful than *tp* and *tas* in all seasons when verified against WFDEI but higher in all lead months against the alternative datasets. The forecast system captures El-Niño Southern Oscillation (ENSO)-related anomalous years with region-dependent skill.

3.1 Introduction

The economies of East Africa and the livelihood of millions of people in the region are heavily affected by climate variability. Advance knowledge of adverse climate anomalies would enhance timely development and implementation of coping mechanisms with respect to food security and water management. Seasonal forecasts may provide such information and, when skillful, have considerable potential to improve the present situation.

Seasonal forecasts provide climate outlooks from a month to over a year ahead based on the interactions of the atmosphere with slowly varying climate system components like the oceans and land surface (Thomson et al., 2006; Kim et al., 2012; Doblas-Reyes et al., 2013; Manzanas et al., 2014). A major source of seasonal predictability is the El-Niño Southern Oscillation (ENSO) because it forces climate anomalies globally (Harrison, 2005; Jan van Oldenborgh et al., 2005; Hoskins, 2006; Palmer, 2006; Charles et al., 2012; Doblas-Reyes et al., 2013; Manzanas et al., 2014) and is predictable with 6–12 months lead-time.

While ENSO originates in the pacific, other tropical ocean basins also exert influence on the climate variability over land. For example, variations in the horizontal structure of tropical Atlantic sea surface temperatures (SSTs) influence the position of the Inter-Tropical Convergence Zone (ITCZ) and consequently the local climate over Northeastern Brazil, West Africa and East Africa (Camberlin et al., 2001; Goddard et al., 2001; Philippon et al., 2002; Semazzi et al., 1988; Diro et al., 2011; Doblas-Reyes et al., 2013).

Rainfall variability over southern and eastern Africa show strong correlation to ENSO (Goddard et al., 2001; Black et al., 2003; Liebmann et al., 2014; Smith and Semazzi, 2014). However, the Indian Ocean Dipole (IOD) is also known to drive climate variability over East Africa, especially during the 'short rains' (OND) (Goddard et al., 2001; Saji and Yamagata, 2003; Black, 2005; Owiti and Ogallo, 2007; Owiti et al., 2008) either acting alone or together with ENSO variability.

Nowadays Global or Regional Climate Models (GCM/RCMs) are used to produce seasonal forecasts. Over the last decade, these dynamical models have improved such that their forecast skill now surpasses that of statistical models in some seasons (Jan van Oldenborgh et al., 2005). Even so, uncertainties related to initial boundary conditions, data assimilation, external forcings and model formulation all limit the skill of a single deterministic forecast. Probabilistic approaches overcome these limitations through ensemble prediction (Kalnay et al., 2006) using multiple realizations for a single forecast time and location to sample forecast uncertainty. Ensemble generation is achieved by either perturbations of initial conditions, perturbations introduced at each model integration (stochastic physics) or use of multi-model ensembles (Graham et al., 2000; Tebaldi et al., 2005; Thomson et al., 2006; Shutts et al., 2011; Weisheimer et al., 2011; Doblas-Reyes et al., 2013; Weisheimer and Palmer, 2014).

This study analyses the performance of one such dynamical forecasting system i.e. the ECMWF Ensemble Prediction System-4 (hereafter *S4*). A number of studies have examined the skill of ECMWF ensemble prediction systems starting with system-1 introduced in 1997 (Stockdale et al., 1998) to the current *S4* (Molteni et al., 2011).*S4* was found to predict El-Niño/La-Niña phenomenon better than statistical models especially in spring season (Jan van Oldenborgh et al., 2005). It has been assessed for skill in predicting Asian summer monsoons (Kim et al., 2012); Northern hemisphere winter (Kim et al., 2012); global meteorological drought (Dutra et al., 2014); below normal rainfall in the horn of Africa (Dutra et al., 2013) and drought forecasting in East Africa (Mwangi et al., 2014). All analysis over east African domain concentrated on the evaluation of precipitation forecasts. Critically needed is evaluation of other agriculturally important variables such as shortwave radiation and temperature. This study first aims to evaluate the prediction skill of *S4* over Eastern Africa for all three agriculturally relevant climate variables.

When used to drive impacts models (e.g. agriculture or hydrology), the performance of raw and bias-corrected driving climate information vary. For example, at climate projection time scales, dynamical downscaling does not improve an agricultural model simulation of maize yields over the United States (Glotter et al., 2014). However, statistically bias-corrected driving data did show improvement over the raw GCM and raw RCM downscaled projections. This is good news for poor countries that cannot invest in RCMs. When forecast and observations are in near-equal spatial resolutions, it is unknown whether dynamically downscaling to a higher resolution than verifying data would be valuable, but bias correction still is. Therefore, the second objective of this study is to assess the value of bias correction of *S4* forecasts over East Africa.

The potential application of such forecasts in agricultural impacts assessments to enable timely adaptation requires predictive skill of the relevant variables and information on associated forecast uncertainty before the respective cropping seasons. Thus, the third objective of the present work is to assess the skill of *S4* for each cropping season relevant to the different regions of East Africa.

Thus, we address the following research questions:

1. How well does the *S4* simulate the climatology and seasonality of the variables relevant to agricultural impacts modelling over East Africa?

2. How does climate forecast skill for these variables at various lead-times compare for each season relevant to different sub-regions?

3. How does climate forecast skill for these variables at various lead-times compare for each season relevant to different sub-regions?

4. Does bias correction either improve or adversely affect skill for lead-times, seasons and geographical forecast units assessed?

5. How well does *S4* simulate anomalous wet and dry years associated with positive or negative ENSO phases?

3.2 Data description

S4 reforecasts starting 1981 to 2010 is evaluated. This is a fully coupled ocean–atmos phere GCM based on the Integrated Forecast System (IFS c36r4) atmospheric model at a resolution of TL255 (approximately 0.75° horizontal resolution) and 91 vertical levels. Forecast initialization starts on the first day of each month with 15-perturbed initial conditions giving 15-ensemble members, each providing a forecast 7 months into the future. A combination of atmospheric singular vectors and an ensemble of ocean analysis provide a means to create the perturbations (Molteni et al., 2011).

Forecast verification requires good observational data (Maraun et al., 2010) and robust methodologies. Over East Africa, a sparse climatological station network limits use of pure in situ observations to verify gridded forecast products. We therefore use a gridded model–observation fusion data product, the Climate Research Unit (CRU) bias-corrected version of the Water and Global Change (WATCH) forcing data ERA-Interim (WFDEI) (Weedon et al., 2010, 2011, 2014; Harding et al., 2011) as reference for precipitation, near-surface air temperature and downward surface shortwave radiation. Since, ERA-Interim and *S4* use the same atmospheric model (though different versions; c31r2 vs c36r4, resp.), independence of the forecast and reference data is inadequate, even though the former assimilates, and is bias corrected to, *in situ* observed data and the latter does not.

Therefore, in addition to the WFDEI, we use separate alternative data sets. The African Rainfall Climatology version 2 (ARC2) developed by the NOAA Climate Prediction Centre for precipitation, radiation data from NASA/GEWEX Surface Radiation Budget release-3.0 data (SRB3), and University of Delaware (UD11) near-surface air temperature data.

ARC2 is a 0.1° resolution daily precipitation data from 1983 to the present derived from a combination of EUMETSAT observed 3-hourly infra-red precipitation estimates and gauge observations from WMO's Global Telecommunication System (Novella and Thiaw, 2013). It is operationally used by the US Agency for International Development's Famine Early Warning Systems Network (USAID-FEWS NET) to generate hazard outlooks over regions of Africa. Note that ARC2 is a real-time product and may not include as many gauge observations as CRU used in WFDEI.

SRB3 is a global 1° dataset available from July 1983 to December 2007 (Mlynczak et al., 2010; Raschke et al., 2006). UD11 is a 0.5° global dataset starting from 1901 to 2014 produced from a combination of the Global Historical Climate Network (GHCN) (Lawrimore et al., 2011) and archives of Legates and Willmott (Willmott and Matsuura, 2001) and http://climate.geog.udel.edu/ climate/html_pages/Global2011/REA DME.GlobalTsT2011.html).

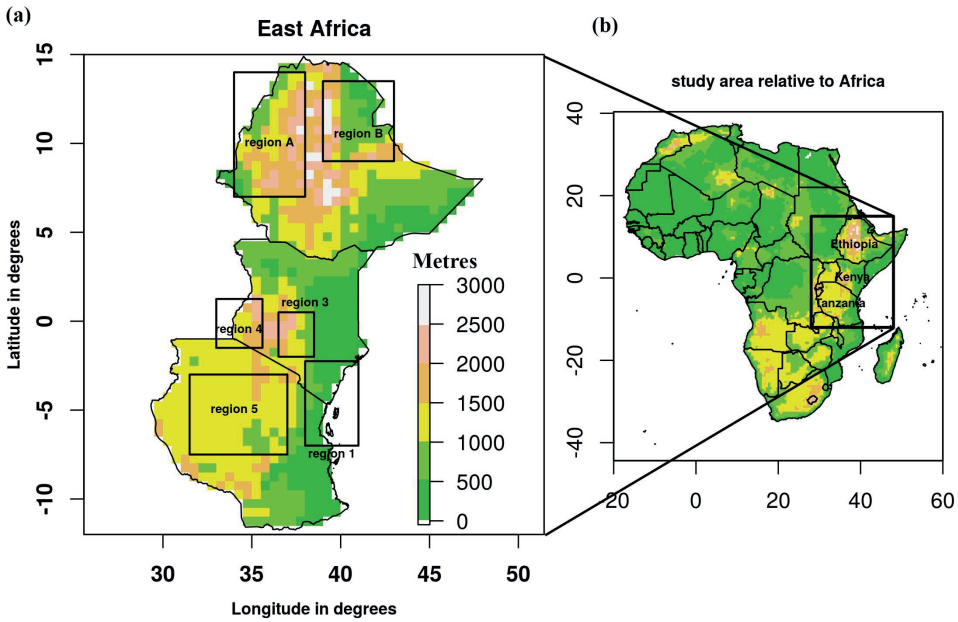


Figure 3.1: The study area and homogeneous rainfall regions (a) relative to African continent (b). Colours show 0.5° gridded elevation (in metres) data downloaded from University of Washington Archive (http://www.jsao.washington.edu/data-climate_data_archive).

3.3 Methodology

We downloaded WFDEI and *S4* precipitation rate (*tp*), near-surface air temperature (*tas*) and surface downwelling shortwave radiation (*rsds*) data for East Africa (Figure 3.1) from the ECOMS User Gateway (Magariño et al., 2014; http://meteo.unican.es/ec oms-udg). We accessed SRB3 data from https://eosweb.larc.nasa.gov/project/srb, ARC2 data from http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.F EWS/.Africa/.DAILY/.ARC2/ and UD11 data from http://www.esrl.noaa.gov/psd/da ta/gridded/data.UDel_AirT_Precip.html.We interpolated all data from their native grids to the 0.5° WFDEI grid using a bilinear method with an appropriate land mask.

Bias correction was performed by empirical quantile–quantile mapping (qqmap) approach (Jakob Themeßl et al., 2011; Amengual et al., 2012; Lafon et al., 2013) on daily forecast data against each of the reference datasets by use of one common set of correction parameters calibrated for each month on the ensemble mean then applied on each of the 15-ensemble members individually.

Forecast aggregation to monthly means enabled skill evaluation through 7 months lead-time. Aggregation into 3-month seasonal means allow evaluation at 0–4 months before start of the relevant rainfall seasons i.e. JJA, MAM and OND seasons in the northern, inner equatorial region and the southern parts of the study area, respectively. Lead-time refers to the number of months the forecast started before a target month or season. We evaluate skill at grid points and for regions of similar rainfall regimes shown in Figure 3.1. Regions A and B have been delineated using principal component analysis on observed annual rainfall in Ethiopia by Eklundh and Pilesjö (1990) (their Figure 7), while the remaining regions are from a combination of empirical orthogonal function and simple correlation analysis on monthly rainfall totals by Indeje et al. (2000) (their Figure 1(b)).

3.3.1 Skill assessment

We use mean error (bias) (Willmott, 1982; Legates and McCabe Jr, 1999; Willmott and Matsuura, 2005; Willmott et al., 2012) to show the mean difference between forecasts and the reference data. Assuming that each ensemble member is an equally probable forecast, conversion to forecast probabilities enable validation by use of two verification measures, the Ranked Probability Skill Score (RPSS) and the Relative Operating Curve Skill Score (ROCSS) for above normal (AN) and below normal (BN), or upper and lower-tercile forecasts, respectively. ROCSS and RPSS compare the skill of a forecast to that of a standard reference (i.e. the climatological forecasts and observed climatology, respectively) such that zero means the forecast is as good as the reference. Positive values imply an improvement, and negative values imply no skill. See Appendices B.1 and B.2 for brief descriptions. We apply Spearman's rank correlation coefficient and its significance to assess the correspondence between average ensemble forecast and the reference data anomalies. R-packages,'SpecsVerification' (Siegert, 2015) and 'easyVerification' (Bhend et al., 2016) are used to calculate the RPSS and ROCSS, respectively with standard error to mark significantly positive skill. We used 'downscaleR' package (https://github.com/SantanderMetGroup/downscaleR) for bias correction and data interpolation.

The ability to capture ENSO associated anomalous rainfall is assessed based on anomalous years identified from NOAA's Oceanic Niño Index (ONI) based on the Extended Reconstructed SST version 4 (ERSST.v4) (Smith et al., 2008; Liu et al., 2015) and available at http://www.cpc.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml. As an example, 1982/83, and 1997/98 represent El-Niño years while 1983/84, 1988/89 and droughts of 1999/2001 represent La-Niña years, agreeing with other studies (i.e.Camberlin and Philippon (2002); Camberlin et al. (2001); Indeje et al. (2000); Meyers et al. (2007); Omondi et al. (2013)).

3.4 Results

3.4.1 Simulated climatology and inter-annual variability

In this section, spatio-temporal patterns of *tp*, *tas* and *rsds* averaged over 15-members in the period from 1981 to 2010 are compared to the reference data sets at each grid point for MAM, JJA and OND seasons. We consider bias (mean error) in Figures 3.2 to 3.4, and mean correlation (and its significance) in Figures 3.5 to 3.7.

3.4.1.1 Precipitation rate (*tp*)

The skill of a model depends on the quality of reference data. The two reference data sets agree better in JJA than in MAM and OND seasons in which differences in mean spatial patterns and biases exist when validated against WFDEI and ARC2 (Figure 3.2). Relative to WFDEI, MAM forecast exhibits dry biases over large swathes of the study area while validation against ARC2 shows lower biases. In southeastern Tanzania, biases of opposing signs (i.e. up to $5\,\mathrm{mm\,d^{-1}}$ against WFDEI and $-2\,\mathrm{mm\,d^{-1}}$ against ARC2, respectively). ARC2 is known to underestimate *tp* over East Africa and especially summer rainfall in Ethiopia (Novella and Thiaw, 2013). This may explain the dominance of *S4* wet bias when validated against ARC2 in JJA and OND. Biases against WFDEI are lower in both JJA and OND, possibly stemming from *S4* initial conditions (also ERA-interim reanalysis).

In MAM, extents of region with a dry bias enlarge with lead-time over western Kenya (verified against both WFDEI and ARC2) and northern Tanzania (against WFDEI).The change in bias with lead-time is unsystematic, getting better with forecasts initialized in the months of November and December (lead-time 4 and 3). This could result from either model drift or due to the stability of ENSO circulation in November–December even though available literature does not establish the significance of ENSO in MAM rainfall (Camberlin and Philippon, 2002). A similar characteristic is evident over Ethiopia in JJA where biases reduce with increasing lead-time.

In OND, dry biases in central Kenya (against WFDEI) and wet biases in southern Tanzania (against ARC2) do not drastically change with forecast lead-time suggesting an existing influence of local features such as surface topography. The ITCZ position in southern Tanzania from October–May influences rainfall and could explain the persistent of *tp* biases.

Figure 3.5 shows good temporal pattern correlation ($0.2 \leq r \leq 0.8$) at lead-0, with significance over large extents of the study area. Beyond lead-0, the relative strengths differ depending on the reference data and lead-time. In JJA (Figure 3.5b, ARC2 show poorer correlation to the forecasts even though it is known to reproduce well the inter-annual variability of precipitation (Novella and Thiaw, 2013). A physical understanding of the JJA patterns require analysis of associated atmospheric and oceanic circulations not undertaken in this study.

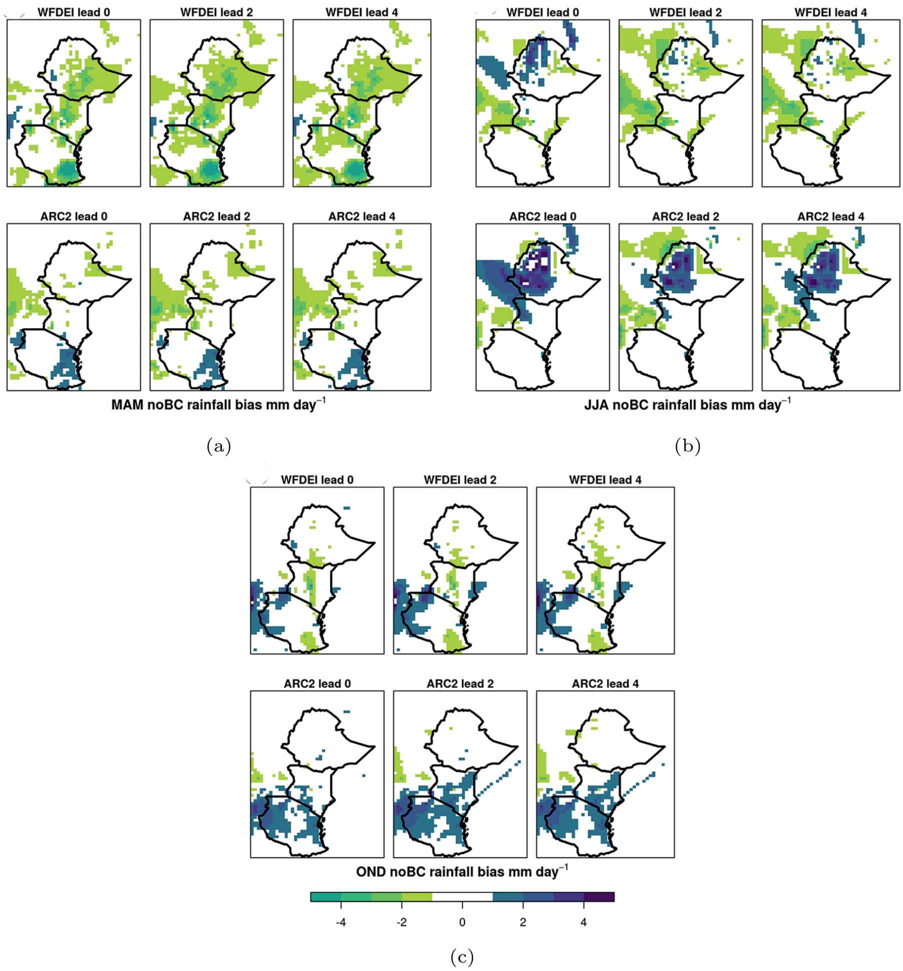
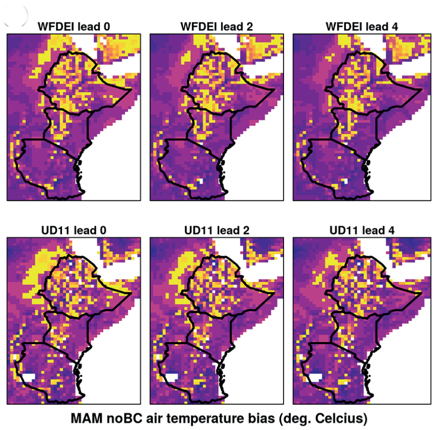


Figure 3.2: Thirty year average mean error in raw *S4* precipitation forecasts (*tp*) compared to WFDEI and ARC2 reference data for MAM (a), JJA (b) and OND (c). Negative (positive) values show wet (dry) biases. Plots are shown for forecast lead-time 0, 2 and 4 months before start of each season.

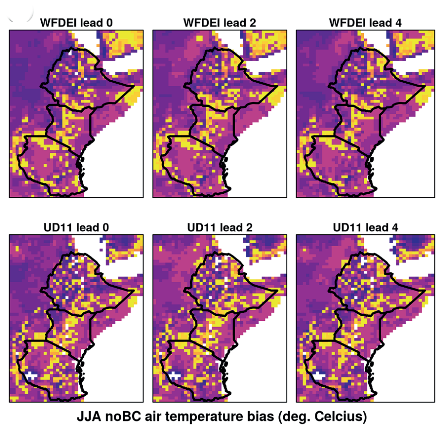
We conclude that *S4* simulates well the inter-annual variability, spatial patterns and structure of *tp* with bias characteristics that are dependent on season and the validating dataset. Similar bias patterns in JJA suggest closeness in precipitation in both reference data sets.

3.4.1.2 Near-surface air temperature (*tas*)

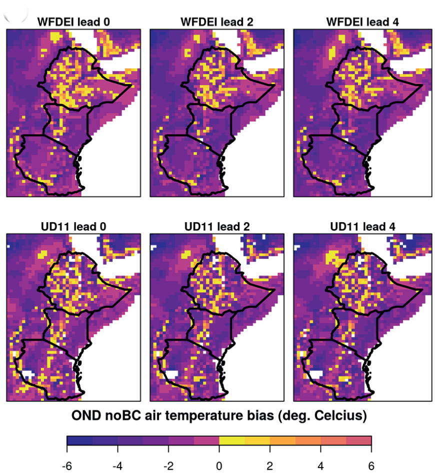
Figure 3.3 shows the spatial fields of mean *tas* biases. While the spatial structure is fairly simulated, cold biases (<3°C in some grid cells) dominate large areas in



(a)



(b)

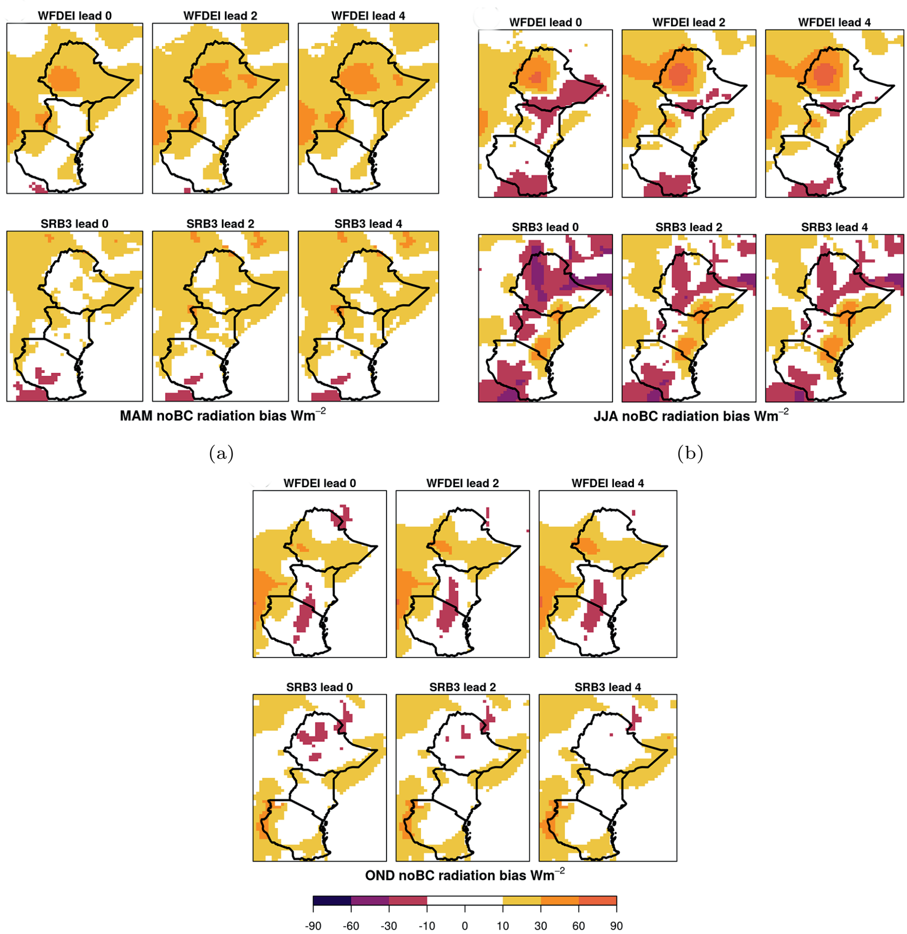


(c)

Figure 3.3: Thirty year average mean error in raw *S4* precipitation forecasts (*tas*) compared to WFDEI and UD11 reference data for MAM (a), JJA (b) and OND (c). Negative (positive) values show cold (warm) biases. Plots are shown for forecast lead-time 0, 2 and 4 months before start of each season.

all seasons irrespective of the reference. Apparent mix of both warm and cold *tas* biases characterize the JJA season. Temperature biases against WFDEI and UD11 are similar especially in JJA, while biases in *tp* (Figure 3.2) and *rsds* (Figure 3.4) are not. Hence, the link between cloudiness and temperature cannot explain this JJA temperature bias. Despite these biases, there is a high correlation of *tas* forecast with both WFDEI and UD11 in JJA and OND over four lead-months (Figure 3.6). The correlation weakens in some grid cells at lead-4 of MAM forecasts but is still positive in many places.

Thus, we conclude that magnitudes, patterns and direction of temperature biases are



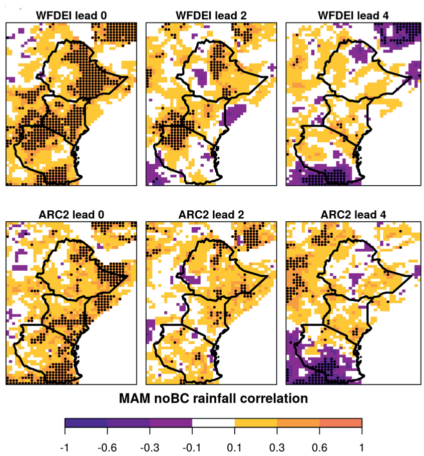
(c)

Figure 3.4: Thirty year average mean error in raw *S4* precipitation forecasts (*rsds*) compared to WFDEI and SRB3 reference data for MAM (a), JJA (b) and OND (c). Negative (positive) values show cold (warm) biases. Plots are shown for forecast lead-time 0, 2 and 4 months before start of each season.

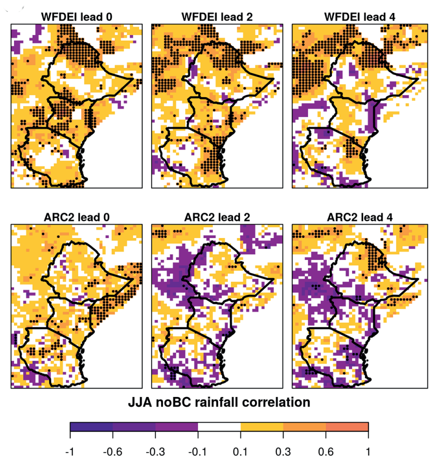
largely similar in the three seasons and nearly constant with lead-time, irrespective of the reference dataset. Temperature biases seem to correlate with elevation, i.e. warm biases in the high grounds upwards of 1500 m elevation (Figure 3.1), and cold biases at lower elevations.

3.4.1.3 Surface downwelling shortwave radiation (*rsds*)

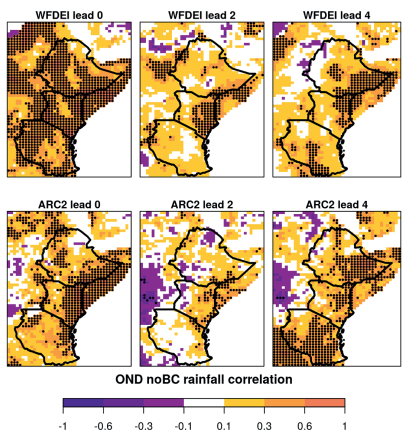
Overestimation of *rsds* against WFDEI and underestimation against SRB3 are apparent over Ethiopian highlands in JJA (Figure 3.4). Regions of high positive bias



(a)



(b)



(c)

Figure 3.5: Thirty year average anomally correlation coefficient between raw *S4* precipitation forecasts (*tp* against WFDEI (top row), ARC2 (bottom row) for MAM (a), JJA (b) and OND (c) seasons. Dots show areas of significant correlation at 95% level. Plots rare shown for forecast lead-times 0, 2 and 4 months before start of each season.

against WFDEI ($\approx$80W m^{-1}) extend with increasing lead-time. Validation against SRB3 shows lower magnitudes of negative bias apparent at lead-0. Low *rsds* amplitudes in WFDEI may result from inherent absence of elevation correction in WFDEI *rsds* data (Weedon et al., 2014). Drier regions in each season also exhibit negative *rsds* biases related to atmospheric conditions such as cloudiness and related properties. In MAM, influence of lead-time on *rsds* bias is not obvious except over an area of positive bias in Ethiopia that enlarges with increasing lead-time. Less bias is apparent in OND relative to other seasons. There is however a region running from southeastern

Tanzania through central to northern Kenya that exhibits negative bias unique to WFDEI validation that seems to correspond with the Rift Valley escarpment.

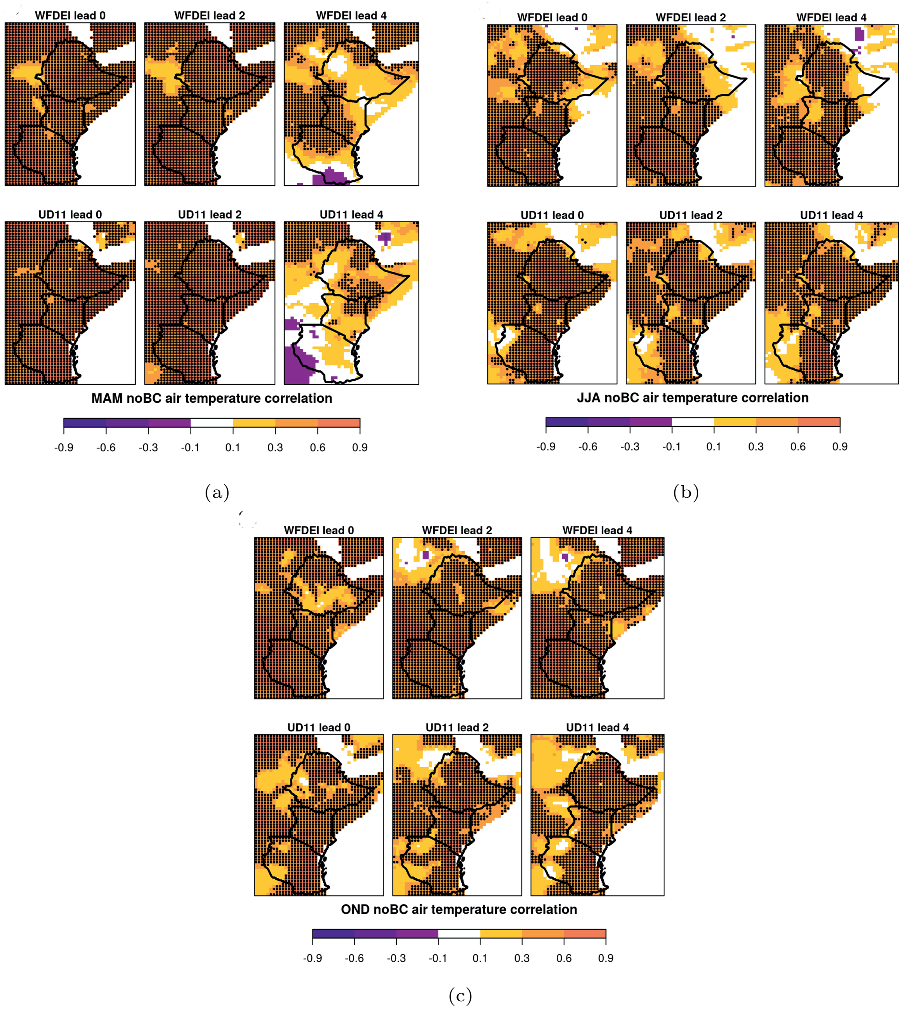
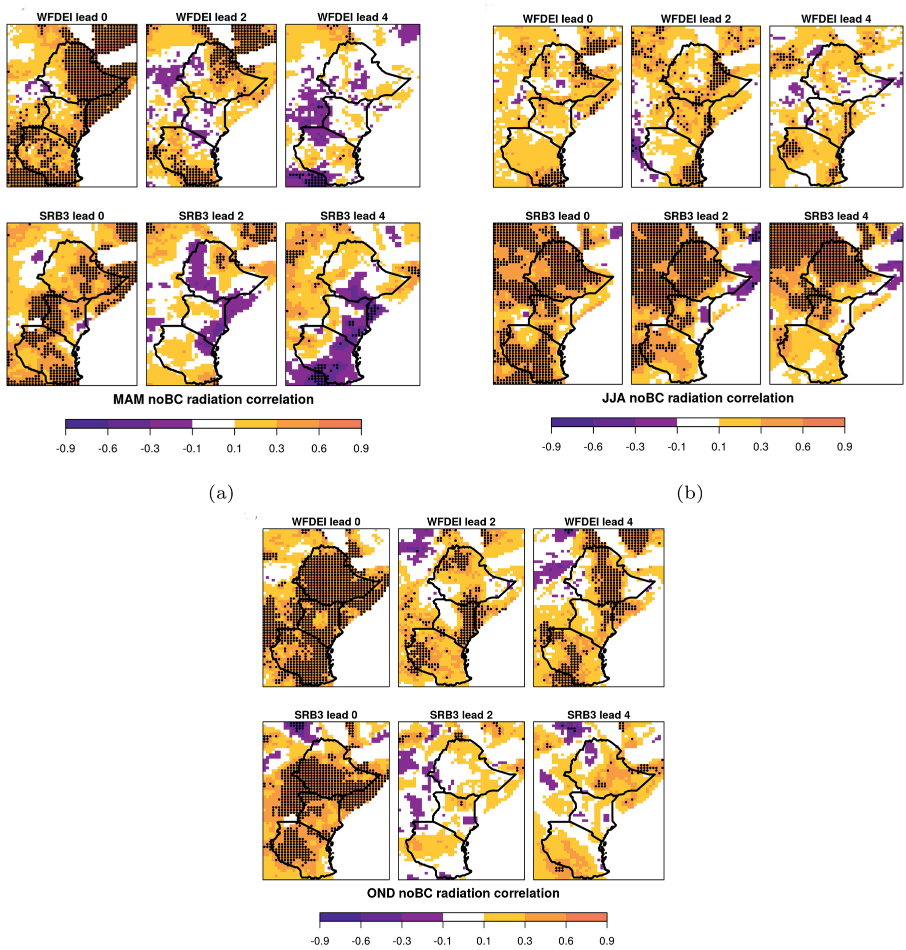


Figure 3.6: Thirty year average anomally correlation coefficient between raw *S4* near-surface air temperature (*tas* against WFDEI (top row), D11 (bottom row) for MAM (a), JJA (b) and OND (c) seasons. Dots show areas of significant correlation at 95% level. Plots rare shown for forecast lead-times 0, 2 and 4 months before start of each season.

Regions of significant correlations in rsds (r≥2) are apparent at lead-0, against both the WFDEI and SRB3 (Figure 3.7). Verified against SRB3, more grids show good and significant correlations in all lead-times in JJA. Against WFDEI, higher correlations and significance in OND exist over all lead-times. The numbers of grid cells with



(c)

Figure 3.7: Thirty year average anomaly correlation coefficient between raw *S4* downward surface shortwave radiation forecasts (*rsds* against WFDEI (top row), SRB3 (bottom row) for MAM (a), JJA (b) and OND (c) seasons. Dots show areas of significant correlation at 95% level. Plots rare shown for forecast lead-times 0, 2 and 4 months before start of each season.

significant correlations reduce with lead-time in all seasons. An understanding of anti-correlations seen in MAM and present for either reference data would require study of cloudiness patterns or regional aerosol load patterns, among others not studied in this paper.

Concluding, the behavior of biases in *rsds* suggests dependence on validation data set, the seasons' atmospheric conditions related to cloudiness and its properties, and on surface conditions related to elevation. High rainfall–high cloudiness areas show a positive radiation bias and a negative bias in the low rainfall areas.

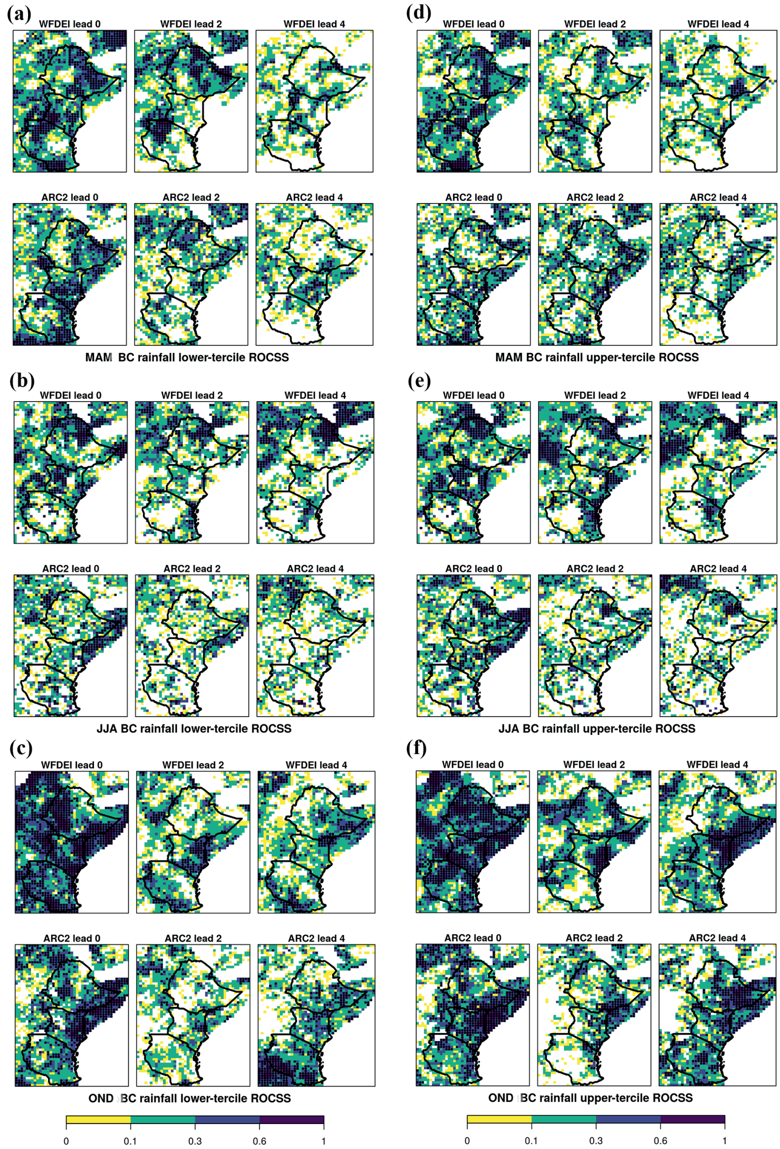


Figure 3.8: Relative operating curve skill score (ROCSS) for lower-tercile (column 1), and upper-tercile (column 2) *S4* bias corrected precipitation (*tp*) forecasts. Figures are shown for MAM, JJA, OND seasons and for lead-times 0, 2 and 4 before start of each season. Only areas of skill (i.e. ROCSS>0) are shown with shades. Dots show areas where ROCSS is significantly greater than zero at 95% level.

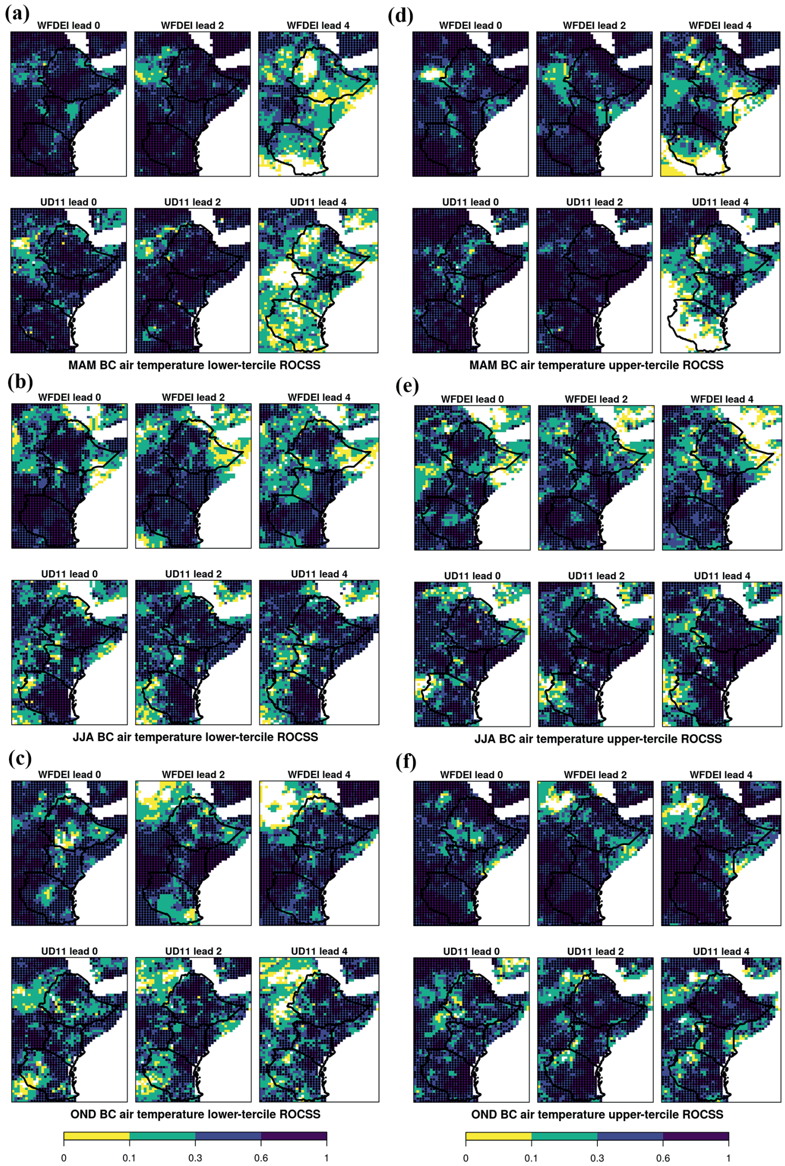


Figure 3.9: Relative operating curve skill score (ROCSS) for lower-tercile (column 1), and upper-tercile (column 2) *S4* bias corrected near-surface air temperature (*tas*) forecasts. Figures are shown for MAM, JJA, OND seasons and for lead-times 0, 2 and 4 before start of each season. Only areas of skill (i.e. ROCSS>0) are shown with shades. Dots show areas where ROCSS is significantly greater than zero at 95% level.

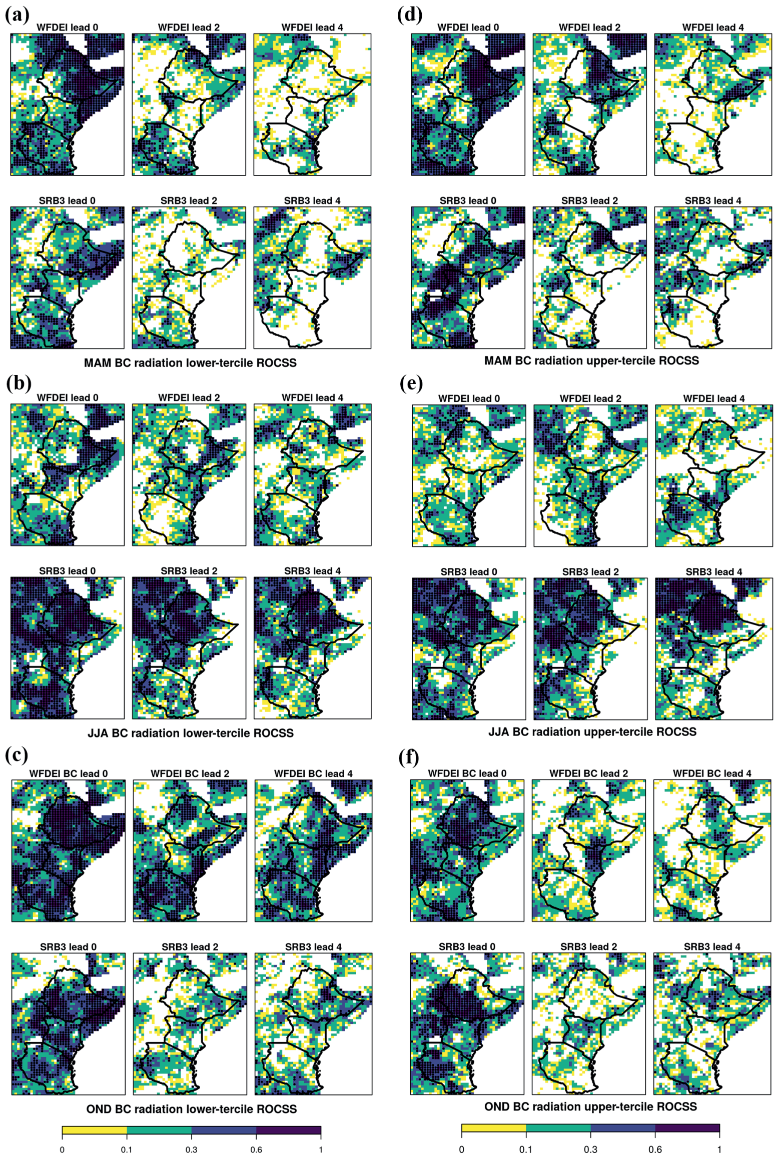


Figure 3.10: Relative operating curve skill score (ROCSS) for lower-tercile (column 1), and upper-tercile (column 2) *S4* bias corrected downward surface shortwave radiation (*rsds*) forecasts. Figures are shown for MAM, JJA, OND seasons and for lead-times 0, 2 and 4 before start of each season. Only areas of skill (i.e. ROCSS>0) are shown with shades. Dots show areas where ROCSS is significantly greater than zero at 95% level.

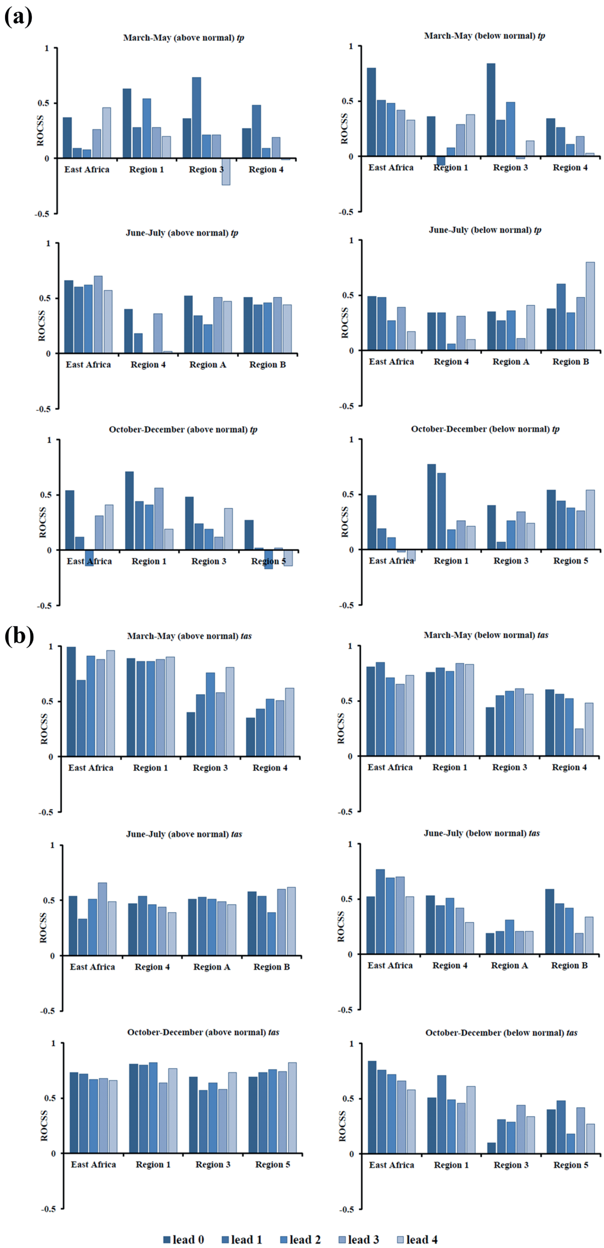


Figure 3.11: ROCSS averaged over homogeneous regions of the study area for MAM, JJA and OND seasons. Bar graphs are for lower-tercile (beloe-normal) and upper-tercile (above-normal) for precipitation (*tp*) (a), and near-surface air temperature (*tas* (b). Results are shown to illustrate the influence of spatial aggregation of skill.

Table 3.1: Precipitation ROCSS for single grid point from each of the homogeneous rainfall regions

	MAM lower-tercile						MAM upper-tercile				
	Lead 0	Lead 1	Lead 2	Lead 3	Lead 4		Lead 0	Lead 1	Lead 2	Lead 3	Lead 4
Region 1	0.39	-0.03	-0.1	0.27	0.405	Region 1	0.21	0.21	0.35	0.4	0.41
Region 3	0.26	0.36	0.115	-0.275	0.2	Region 3	0.53	0.59	0.16	0.07	0.39
Region 4	0.415	0.335	0.255	-0.075	-0.025	Region 4	0.19	0.61	0.09	-0.12	0.06
Region 5	0.125	0.025	0.03	-0.015	-0.02	Region 5	0.08	0.05	-0.02	0.11	-0.31
Region A	0.15	0.485	0.28	0.055	0.125	Region A	-0.32	0.15	-0.06	0.21	-0.10
Region B	0.585	0.46	0.34	0.255	0.3	Region B	0.33	0.66	0.37	0.16	-0.29
	JJA lower-tercile						JJA upper-tercile				
	Lead 0	Lead 1	Lead 2	Lead 3	Lead 4		Lead 0	Lead 1	Lead 2	Lead 3	Lead 4
Region 1	0.2	0.32	0.51	0.55	0.18	Region 1	0.18	0.4	0.54	0.34	0.28
Region 3	0.58	0.35	0.24	0.29	0.08	Region 3	0.4	0.35	0.36	0.14	0.16
Region 4	0.02	-0.10	-0.07	-0.26	-0.13	Region 4	-0.14	-0.13	0.14	-0.32	-0.25
Region A	0.29	-0.23	-0.21	-0.06	-0.14	Region A	-0.05	-0.20	-0.33	-0.26	-0.14
Region B	0.5	0.55	0.62	0.48	0.62	Region B	0.44	0.4	0.4	0.06	0.23
	OND lower-tercile						OND upper-tercile				
	Lead 0	Lead 1	Lead 2	Lead 3	Lead 4		Lead 0	Lead 1	Lead 2	Lead 3	Lead 4
Region 1	0.615	0.665	0.38	0.015	0.32	Region 1	0.6	0.13	0.37	0.65	0.405
Region 3	0.28	0.305	0.52	0.235	0.12	Region 3	0.43	0.135	0.145	0.265	0.175
Region 4	0.58	0.065	0.36	0.155	0.385	Region 4	0.375	0.17	0.24	0.21	0.25
Region 5	0.39	0.15	0.16	0.22	0.35	Region 5	0.33	0.08	-0.13	-0.04	0.11
Region A	0.565	0.4	-0.04	-0.07	0.28	Region A	0.65	0.025	0.25	0.29	0.155
Region B	0.3	0.11	0.12	0.045	0.005	Region B	0.595	-0.06	0.28	0.295	0.025

Number in bold indicate ROCSS that are significantly larger than zero at 95% confidence level. Selected grid points are; Region 1 (Lon. = 39.5° ,Lat. = -4.25°); Region 3 (Lon.=37.25°, Lat.=-0.75°); Region 4 (Lon. = 34.74°, Lat. = -0.25°); Region 5(Lon. = 33.75°, Lat. = -7.75°); Region A(Lon. = 35.25°, Lat. = 9.75°); Region B(Lon. = 39.75°, Lat. = 11.205°.

3.4.2 Grid point RPSS and ROCSS

Inadequate skill in middle-tercile forecasts negatively affects RPSS, suggesting a lower potential usefulness than actually exists. We therefore concentrate on ROCSS and refer to Figures C.1, C.2, and C.3 in the Supporting information for grid point *tp*, *tas* and *rsds* RPSS, respectively. Figures 3.8, 3.9, and 3.10 show grid point *tp*, *tas* and *rsds* ROCSS for MAM, JJA and OND seasons' lower- and upper-tercile forecasts, respectively. Verified against both WFDEI and ARC2, lower- and upper-tercile *tp* forecasts are better than the climatological forecasts by 20–80% in MAM season in many grid points up to at least lead-2. In JJA, lower- and upper-tercile skill of 20–80% encompasses the entire study area through lead 0–4 against WFDEI and less against ARC2 even though the patterns remain similar. High OND precipitation forecast skill at lead-0 depreciates with increasing lead-time but with different rates that are region dependent. In general, MAM upper and lower-tercile *tp* forecasts possess lower skill than in both JJA and OND.

The upper- and lower-tercile *tas* forecast skill (Figure 3.9) are 40–100% better than climatology in all seasons up to at least 3-months before start of season but decay with lead-time. OND and JJA forecasts are skillful in all the four lead-months. Figure 3.10 shows lower- and upper-tercile ROCSS for *rsds* forecasts for all seasons. In MAM, upper and lower-terciles show skill in many grid points (up to 60% better than climatology) irrespective of the reference dataset. Even though there is loss of skill with increasing lead-time, usable skill exists at lead-2 in the northern part. Verification against WFDEI exhibits lower skill in JJA than against SRB3 (skillful up to lead-4) especially in the northern parts of East Africa. In OND, lead-0 forecasts show higher skill for both lower- and upper-terciles irrespective of the reference datasets. Verified against WFDEI however, lower-tercile forecast shows skill from lead-0 through to lead-4 (Figure 3.13c. The lower-terciles are predictable with good skill in all lead-times while the upper-tercile possess skill at lead-0 and 1 only beyond which the forecasts are generally as good as, or only marginally better than climatological forecasts.

In conclusion, grid-point analysis shows that forecast skill of *tp* and *rsds* depend on season, lead-time and the choice of verifying dataset. Choice of reference data set is irrelevant for temperature. In all instances, the middle-terciles forecasts are either as good as the climatological forecasts or worse. Bias correction does not result in significant improvement in grid point ROCSS (see Figures C.4, C.6, and C.7 for raw model simulations of *tp*, *tas* and *rsds*, respectively).

3.4.3 Regional ROCSS

We use *tp* and *tas* forecasts plus WFDEI to show variation of regionally averaged skill. Bar graphs compare ROCSS averaged over East Africa to the scores averaged over the various homogeneous rainfall regions for *tp* (Figure 3.11a and *tas* (Figure 3.11b. There generally is a regional dependency of *tp* ROCSS in all the seasons, terciles and lead months. Noteworthy are MAM and OND above-normal forecasts in which the skill averaged over East Africa is less than for each of the sub-regions. Table 3.1 shows precipitation ROCSS for single grid points from each homogeneous rainfall regions. Results show that good skill at grid point will not necessarily result in good skill over a region (and vice versa). An example is JJA lower-tercile *tp* forecasts in region-4 with apparent strong positive scores at the regional level but negative in a single grid. Above-normal *tas* forecast ROCSS (Figure 3.11b) are better than climatological forecasts irrespective of the spatial extent of aggregation. Below-normal ROCSS is also dependent on region and lead-time. Monthly ROCSS are season, forecast tercile, and lead-time dependent. Good forecast skill at monthly timescales especially at the start of respective seasons (i.e. March, July and October) implies that the start of seasonal rainfall may be correctly simulated and is quite important for application purposes, for example in determining proper crop planting dates in agricultural impact assessments.

3.4.4 Prediction of particular anomalous years and seasons

Figures 3.12, 3.13, and 3.14 (companion figures for non-bias corrected forecasts are in Figures C.7 to C.9) show probabilities assigned to each predicted bias-corrected precipitation tercile (colour bar) and the tercile in which observations fell (small circles) for the period of 1981 to 2010 for MAM, OND and JJA (regions A and B only), respectively. By examining the observation position and the forecast probabilities, it is possible to assess the year-to-year performance and hence anomalous dry and wet years (similar analysis is done in Manzanas et al. (2014)). We only consider precipitation forecasts. *S4* forecasts do not capture all the dry/wet years whether over East Africa or the sub-regions, see e.g. the above-normal (upper-tercile) rainfall in MAM and OND seasons of 1999/2000 La-Niña years. These are fairly simulated in JJA over Ethiopia (Figure 3.14) i.e. coincides with higher than normal rainfall. Dry conditions of 1983/1984, 1988, 2000/2001 are well simulated. Dissimilar skill scores are evident when analyzed over the sub-regions e.g. in region-1 the model fails to simulate observed 1983/1984 drought conditions. Upper-tercile forecast skill depends on region though the model captures wet conditions of 1982/1983 and 1997/1998. These two very wet periods were very strong Indian Ocean SSTs in addition to the Pacific SSTs (Latif et al., 1999; Black, 2005; Owiti et al., 2008).

El-Niño conditions result in higher global temperatures because of heat transfer from oceans to the atmosphere while La-Niña conditions do the opposite (WMO, 2014). Considering *tas* forecasts of the same anomalous years, (figure not shown), the influence ENSO on *tas* is not obvious in 1997/1998 but captured well in most other years. The later part of the century shows high forecast probabilities of upper-tercile temperatures and observations as opposed to the start of the study period probably related to greenhouse gas driven warming trends in the global atmosphere, hence resulting in higher skill scores.

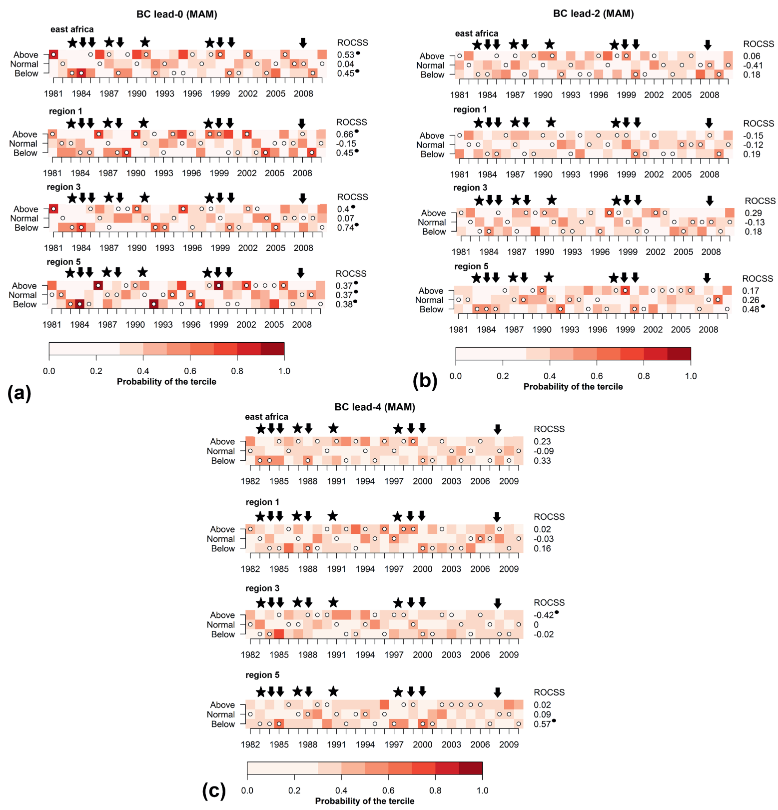


Figure 3.12: Year-to-year bias corrected precipitation textittp forecast probabilities (shades), tercile of occurrence of observation (unfilled circles) and ROCSS for the study period over East Africa and sub-regions in MAM season.Asterisk indicate El-Niño years; arrows indicate La-Niña years and circular dots indicate significant score at 95% level.

3.5 Discussion

3.5.1 Methodology

3.5.1.1 Reference data

Verification of any forecast product preferably uses high quality (*in situ*) observations against which to assess systematic errors, deterministic and probabilistic forecast quality. Current gridded observational climate data practically assimilate station, satellite and model-output data. We used a variety of reference data, each having its strengths and weaknesses, and vary in their degree of independency from each other, and from our forecast product.

Our forecast product *S4* is based on the ECMWF IFS (c36r4) where the hindcasts

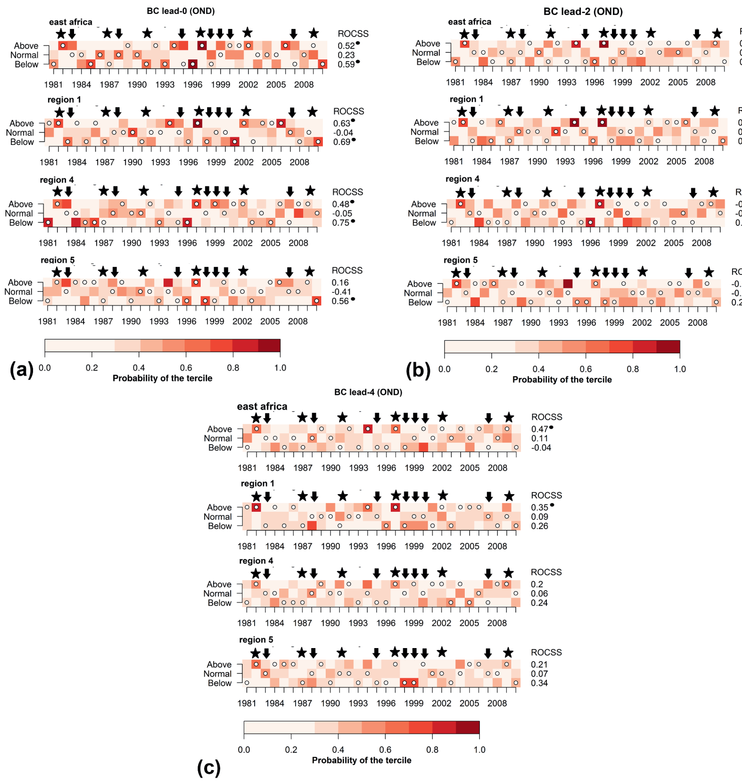


Figure 3.13: Year-to-year bias corrected precipitation textittp forecast probabilities (shades), tercile of occurrence of observation (unfilled circles) and ROCSS for the study period over East Africa and sub-regions in OND season.Asterisk indicate El-Niño years; arrows indicate La-Niña years and circular dots indicate significant score at 95% level.

are each *unconstrained* runs, started from ERA-Interim initial conditions, slightly perturbed to create an ensemble. Our prime reference is the WFDEI, a reanalysis based product in which observations play a role twice. First, during the production of ERA-Interim, the model (also IFS, but c31r32) assimilates observed data (except precipitation) from WMO ground, sea and upper-air networks (Dee et al., 2011), and the model thus recreates a series of constrained daily best possible known status of the climate at the model resolution. Even though the model state is kept as close as possible to the ingested observations, drift and chaotic (convective) processes still produce simulation biases. For WFDEI, these biases are then corrected in a second step against in situ observations from CRU-TS3.1 (Weedon et al., 2010, 2011, 2014), a statistically homogenized, exclusively station-based data set (Harris et al., 2014).

The models used in forecast and reanalysis are largely identical, justifying some questions as to their independency. On the other hand, the unconstrained nature of the forecast runs generates drift characteristics very different from the reanalysis runs. Bias corrections of the latter make the results even more divergent. Therefore, we also used alternative, arguably more independent datasets for the variables under study. For *tp* we used ARC2, a satellite product, but bias corrected, in this case again against WMO-GTS. That makes it independent from WFDEI in its fine scale spatial patterns and daily values, but when aggregated to monthly values, the respective bias corrections will make them more similar again. ARC2 is known to capture inter-annual variability well, but known weaknesses occur in regions of complex topography and rain shadows (Dinku et al., 2007; Novella and Thiaw, 2013), related to interactions between cloud top temperatures and orographic lift and misinterpreted surface temperatures over snow and ice at high elevations. WFDEI precipitation poorly reproduces observed precipitation rates on sub-daily and daily time scales in regions dominated by local convective processes (due to convective parameterisation issues). However, the WFDEI algorithms adjust the monthly precipitation to match the observed numbers of wet days and gauge totals (Weedon et al., 2014). Such reference data peculiarities largely explain the different biases we found over e.g. the Ethiopian highlands versus ARC2 and WFDEI, respectively and it explains the anti-correlations in ARC2 JJArainfall relative to *S4*.

For *tas* we used UD11, a pure station-based product. We do not know the exact degree of overlap in the station network of GHCN (used for UD11) WMO-GTS (for ERA-Interim) and CRU (for WFDEI), but it will be large (GHCN is ingested into both CRU and UD11 datasets) (Tanarhte et al., 2012) and thus their level of dependency. Their statistical homogenization procedures are developed independently, decreasing their dependency.

For *rsds* we used SRB3, a pure satellite-based product.As such it can be assumed to be completely independent from WFDEI and *S4*. SRB3 is known to perform very well against in situ observations (Zhang et al., 2013). WFDEI *rsds* suffers for similar reasons as for *tp* i.e. from poor sub-daily performance under local convective conditions and has minor issues due to lack of elevation corrections (Weedon et al., 2014).

3.5.1.2 Bias correction and skill metrics

Bias correction is essential prior to application of GCM output in impact models, whether on seasonal or longer time scales. For example, a number of thermal time parameters influence crop phenological stages. Any bias in temperature, such as in the *S4* (biases of up to ±4.5°C) will quickly accumulate in thermal times and lead to anomalous simulation of crop development and especially yield fractions. Similarly, systematic biases in *tp* may result in wrong soil moisture status and together with those in *rsds* lead to anomalous overall plant productivity (Liu et al., 2014; Macadam et al., 2016).

For bias correction, we used a quantile–quantile mapping approach (Jakob Themeßl et al., 2011; Amengual et al., 2012; Lafon et al., 2013) on daily forecasts. Alternatives exist but we use qqmap because it can be applied to any variable with no prior assumption of distribution of variables.

Bias correction matches the forecast and observed distribution; nevertheless, in our study it does not significantly improve probabilistic forecast skills. The reasons for this may be twofold. First, the influence of near equality of *S4* and observed data resolutions that averages out the extreme values in any variable, and second, probabilistic validation looks at a range of values in the forecast, typically percentiles. When biases are similar across percentiles, correction will not affect probabilistic metrics especially when these are computed for forecast and reference each in its own distribution. Finally, the relatedness between *S4* and WFDEI, (both depend on very similar model codes and both use ERA-Interim) may also result in a good distribution match. The success of empirical mapping for bias correction using other datasets is because the technique applied acts to match individual observed and simulated distributions without preserving the relationships between variables.

3.5.1.3 Aggregation issues

There exist variations in forecast skill depending on the domain of assessment (i.e. whether assessed over East Africa, in the sub-regions or at a single grid point). This again advocates the importance of skill evaluation at smaller spatial scales since it is already known that rainfall in the region is quite complex and is modulated by many other processes apart from ENSO circulation.Getting an average skill score over the whole domain may not be meaningful to those affected by seasonal climate variability in the sub-regions (the homogeneous rainfall units in this case).

3.5.2 Verification results

3.5.2.1 Model errors

S4, WFDEI, ARC2, UD11 and SRB3 all show the basic pattern and climatology of the region as described in other studies. The shift in rainfall seasons largely follows the ITCZ, but interacts with complex topography, presence of large lakes, variations in vegetation and land-ocean contrasts (Indeje et al., 2000). Northern parts (Ethiopia) exhibit a unimodal rainfall patterns with one season from June to September. Large parts of Kenya and northern Tanzania experience a bimodal pattern receiving its rainfall in MAM, and OND, while western Kenya and the coastal regions exhibit a tri-modal pattern experiencing rainfall in MAM, JJA and OND seasons (Nicholson, 2014; Indeje et al., 2000; Korecha and Barnston, 2007; Gitau et al., 2015).

Errors exist in *S4* that have unique seasonal magnitudes and characteristics determined by the validating data sets. Forecast lead-time generally does not severely alter bias characteristics thereby implying the presence of similar influences.

Bias characteristics in each season may be due to rainfall causing mechanisms associated with regional and local features as opposed to large-scale features that influence OND and JJA seasons (Nicholson, 2014; Camberlin et al., 2001; Camberlin and Philippon, 2002; Ogwang et al., 2014). These could be the influence of water bodies, topography or other meso-scale forcings and their interaction with large-scale mechanisms (Indeje et al., 2000).

Cold bias dominates *tas* simulation, with a persistent structure over lead-time perhaps because local influences such as land-use cover, soil moisture and orography are not well represented in the model resulting in misses when compared to observations. Biases in tas may also depend on the number of station observations included in WFDEI and UD11. The influence of topography on temperature is a feature already identified in other studies such as Ogwang et al. (2014) and exhibited in *S4* lower (higher) temperatures in higher (lower) elevations.

Biases in *rsds* seem to correlate with cloud/rain patterns. Cloud processes remain a challenge in climate model simulations and in *S4*, parameterized cloud/convection and radiation processes explain most of it biases in *tp* and *rsds* (Rotach and Zardi, 2007; Bechtold et al., 2008; Morcrette et al., 2008,?; Jung et al., 2010). To our knowledge this is the first study analysing *rsds* performance in *S4*.

3.5.2.2 Overall forecast skill

Higher precipitation skill (RPSS) seen in OND compared to lower skills in MAM and JJA result from influence of large-scale circulations such as the ENSO and SSTs. Similar results are seen in previous studies such as Hastenrath (1995); Kumar et al. (1999); Mutai and Ward (2000) and Nicholson (2014). There exist less predictability of MAM rainfall because of the influence of local features and processes not well simulated by models. Such influences are discussed in the previous section. JJA rainfall is influenced by other factors such as SST anomalies off the coast of Africa between South Africa and Madagascar (Smith and Semazzi, 2014) accounting for around 51% of the seasonal rainfall variability. This should give the season some predictability though primarily, not only ENSO circulations and local factors influence JJA rainfall but also the Atlantic and Indian Ocean SSTs as illustrated in Nicholson (2014).

Good simulation of rainfall initiated at 3 and 4 months before start of MAM season [i.e. December and November of year $(x - 1)$] can be explained by the maturity of ENSO signal resulting in higher signal-to-noise ratio hence a better predictability before it starts dissipating in the boreal spring season. Available literature does not however establish a significance of ENSO acting alone in MAM seasonal rainfall (Camberlin and Philippon, 2002).

Shortwave radiation forecasts are no better than the climatology in JJA at all forecast lead months when validated against WFDEI, but shows better skill against SRB3. This may result from weaknesses of WFDEI discussed before. Note that RPSS is a summary measure of skill over all terciles and would nevertheless result in low skill even if the model simulates one tercile well.

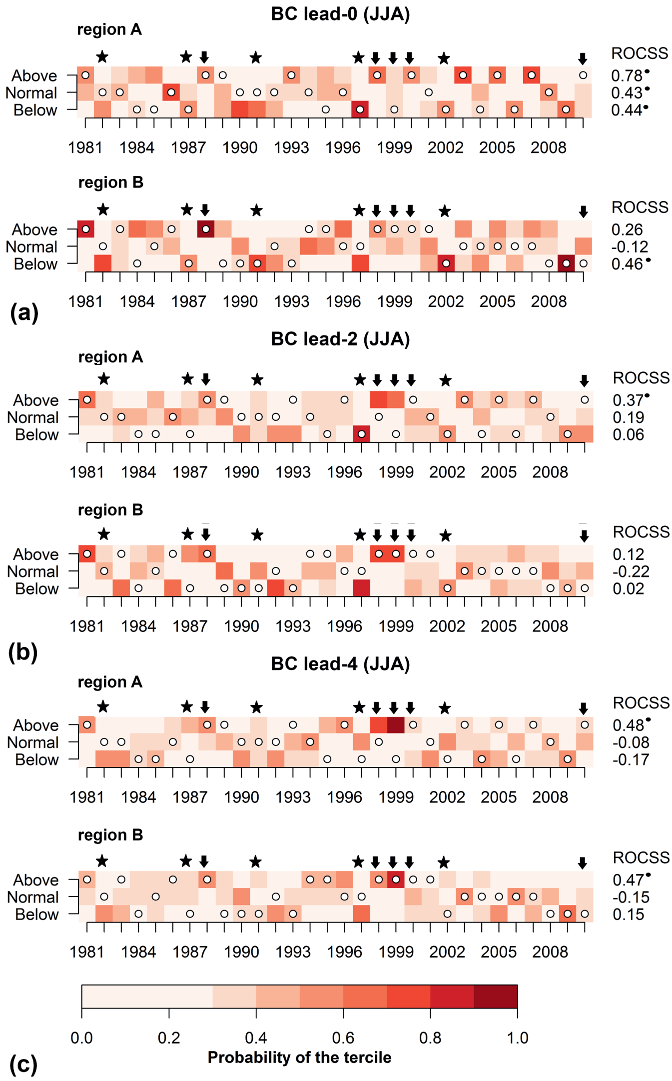


Figure 3.14: Year-to-year precipitation forecast probabilities (shades), tercile of occurrence of observation (unfilled circles), and ROCSS for the study period over sub-regions that receive rainfall in JJA season. Asterisk indicate El-Niño years; arrows indicate La-Niña year and circular dots indicate significant scores at 95% level.

3.5.3 Skill of the seasonal forecasts as a function of geographic forecast units

Another measure, the ROCSS evaluates the performance of forecast terciles. The patterns in skill levels are clearly visible when assessed over East Africa. Differences exist at regional scales perhaps because of the model's inability to simulate small-scale processes. Overall, lower-tercile and upper-tercile forecasts are skillful over many regions compared to commonly low middle-tercile also seen in other studies such as Kharin and Zwiers (2003) and van Van Den Dool and Toth (1991). Lead-time of skillful forecasts depend on regions and seasons reiterating the importance of a detailed validation process in relevant spatio-temporal scales. Verification dataset play a fundamental role in the patterns and magnitudes of ROCSS, but in general, the model fairly simulates above-normal and below-normal terciles compared to near-normal forecasts.

3.5.4 Model simulation of anomalous wet and dry years

Skill in simulation of wet and dry anomalies influenced by El-Niño/La-Niña conditions in the equatorial pacific oceans vary among the sub-regions. Since system-4 simulates well the ENSO conditions (Molteni et al., 2011); forecasts should therefore capture related anomalies over East Africa. Strong ENSO years such as 1982/1983,1997/1998 and dry years of 1984/1985 and 1999, 2000,2001 are captured in most of the regions but missed in others perhaps because ENSO is not the only condition that influence climate variability. It is known that other anomalies such as the IOD (Goddard et al., 2001; Saji and Yamagata, 2003; Black, 2005; Owiti and Ogallo, 2007; Owiti et al., 2008) and the East African lakes exert influence on the local climates (Song et al., 2004; Thiery et al., 2015).

JJA forecasts in the northern part of the study area show good skill in upper-tercile forecasts beyond lead-2 in this study. The GCM provides this advantage over statistical methods. For example, Nicholson (2014) and Korecha and Barnston (2007) found no more than 2 months lead-time in prediction of summer rainfall (JJA) in the GHA. The same studies found shorter lead-times in MAM prediction. The skills shown for MAM forecasts at lead-3 and -4 provide an advantage of the GCM even though lead-1 and -2 also show useful skill in many grid cells. Nicholson (2014), Korecha and Barnston (2007) and Jan van Oldenborgh et al. (2005) argue that spring predictability barrier limited prediction of MAM and summer rains. Good skill in OND precipitation forecast at long lead-times is comparable to other studies employing statistical methods (Philippon et al., 2002; Nicholson, 2014). However, GCM forecasts provide advantage because they are based on physical relationships and are expected to hold into the future unlike the statistical models that may break with future climate change.

This study shows the potential use of *S4* for impact studies. Skill in *tp* is not only important for agricultural impact studies but also hydrological impacts such as surface runoff, river discharge and hydropower generation. It potentially enables estimation of water demand, water availability and water quality months in advance. For example, Region-3 (Figure 3.1) that hosts several hydropower dams show *tp* skill at least 3-months before MAM and OND seasons. Application to hydrological flow simulations would be of high importance to the power generation companies. Agricultural regions over Ethiopia (regions A and B) for example have its cropping season in JJAS. A season in which *tp* and *tas* are potentially predictable with longer lead-times could influence farm management practices to avoid losses several months before start of season. A continuation of this research will assess whether the use of a rather skillful *S4* results in a likewise skillful impact modelling.

3.6 Conclusions

We investigated the potential usefulness of ECMWF ensemble forecasts for East Africa in impacts modelling. For utilization, the forecasted variables must have skill. We therefore validated forecasted precipitation, air temperature and shortwave radiation against WFDEI and other independent climate data sets (i.e. ARC2, SRB3 and DU11 for *tp*, *tas* and *rsds*, respectively) for the relevant seasons, at grid point and sub-regional levels. Deterministic measures; the mean and bias, and probabilistic measures; the RPSS and the ROCSS are considered as metrics of forecast skill. It is shown in general that skill of the three variables (*tp*, *tas* and *rsds*) depends on season, forecast lead-time and location. The role of the quality of a validating dataset cannot be over-emphasized.

S4 realistically reproduces the spatial patterns of observed seasonal climatology, of *tp*, *tas* and *rsds* irrespective of the reference data sets considered but with biases. Precipitation bias characteristics are season-dependent. The error magnitudes depend on verifying data set. Dry precipitation biases characterize the MAM season; wet biases characterize OND and a mixture of both in JJA. Cold biases dominate *tas* forecasts in all seasons, the spatial patterns of which do not show a strong seasonal variation. Warm biases exist over dry regions during each of the three individual seasons. Biases and patterns in *rsds* follow rainfall seasons and the surface topography.

Bias correction of the raw data does not improve probabilistic forecast skill but is important for impact applications. For example, biases of driving temperature or rainfall fields may have gross influences in crop growth or hydrological flow simulations.

Grid point verification using RPSS show that precipitation and shortwave radiation forecasts initialized at the start of the first month show skill only at start of season (lead-0) but skill levels are low. Even though lead-0 forecast skill may be influenced by initial conditions, it is important for impacts assessment and especially in agriculture as it determines the sowing dates (i.e. rainfall onset dates).

The ROCSS reveals that *S4* skillfully discriminates between upper-tercile, middle-tercile and lower-tercile forecasts. Aggregated over East Africa, upper-tercile and lower-tercile precipitation forecasts initialized up to at least three lead-months show skill. Temperature forecasts initialized up to 4 months before start of season are skillful and useful while *rsds* show less skill though still usable. Middle-tercile predictions remain poor for all variables at all seasons and lead-times. In the sub-regions, tercile skills and lead-times of useful forecasts are season and region specific, no general statement is possible.

Verification of yearly OND and MAM forecasts indicate good simulation of anomalous years when assessed over the whole of East Africa. The forecasts capture manifestations of anomalous years in terms of rainfall and temperatures in certain sub-regions (the geographical forecast units) in particular years and seasons. This reiterates the importance of verifying skill at smaller relevant spatial extents relevant for impact assessments but not at larger regions such as East Africa where prediction skill is not uniform.

S4 fairly captures the inter-annual climate variability of the three variables. Inter-annual climate variability affects socio-economic activities (such as food production and hydropower generation) of large populations over East Africa. This is true especially when socio-economic activities depend on highly variable climate parameters like rainfall. When used in impacts models, an assumption is that the skill in driving variables will propagate into the impact models. In this study, skill shown by *S4* grid point, sub-regions and seasons for the three variables suggest a potential application in impact models. We have shown the importance of using more verification measures since a single one cannot capture all forecast attributes.

A follow-up research to this will therefore investigate the influence of precipitation, surface temperature and shortwave radiation in crop production and hydrological simulations over East Africa.

Acknowledgments

This study was financially supported by the EUPORIAS project (EUropean Provision of Regional Impact Assessment on Seasonal-to-decadal timescale); grant agreement No. 308291, funded by the European Commission (EU) project in the Seventh Framework Programme. We confirm that the authors have no conflict of interest to declare. We also like to thank the valued suggestions and insightful comments from two anonymous reviewers that led to an improved version of the manuscript.

Chapter 4

Sensitivity and Predictability of Rain-fed Maize Yields to Growing Season Climate Indices: East Africa

Abstract

This paper explores the influence of rainfall and temperatures indicators on maize yields and their seasonal predictability in northern Ethiopia (North-Gonder) and equatorial Kenya (Bungoma). We use characteristics that define rainfall distribution i.e. evenness (E_R), unevenness (U_R), time of rainfall in a period (Ad). Temperature characteristics including Growing Degree Days (GDD) and Killing Degree Days (KDD). We explore their influence in the entire maize growing season, vegetative, around anthesis, and in the reproductive stages. Reference rain fed yields are obtained from WOFOST crop model driven by WATCH forcing data ERA-Interim for the period 1980-2011; reference weather indicators are derived from the same. We derive predicted ensemble weather indicators from an ensemble of fifteen seasonal climate predictions from the ECMWF's System-4 (*S4*) initialized three forecast lead months before planting dates. Findings show that the relationship between rainfall, temperatures and yields are location specific. While both rainfall and temperature characteristics influence yields in northern Ethiopia, only temperature characteristics are of significant influence in western Kenya. Sensitivity of yields to climate indicators vary with maize phenological stages. Some indicators are predictable in one growth phase, but not in another. In northern Ethiopia, above normal temperature forecasts are more predictable than below–normal forecasts. GDD forecasts are no better than

This chapter is submitted for publication as:
Ogutu G.E.,I. Supit,P. Omondi and R.W. Hutjes, 2020: Sensitivity and Predictability of Rain-fed Maize Yields to Growing Season Climate Indices: East Africa. Journal Climate Services

climatology (i.e. ROCSS<0.3). Forecasts of rainfall indicators during entire growing season and phenological phases are predictable with certainty at lead times–0 and –1. Forecasts of temperature indicators show uncertainty except GDD and KDD but varying with lead-time and growth stage. In Bungoma, growing season rainfall characteristics lack predictability. In the vegetative stage, temperature characteristics are predictable except GDD. Around anthesis, only the minimum and daily temperatures above–normal forecasts are predictable. Forecasts of above–normal maximum, minimum and daily temperatures are predictable in the reproductive stage. The results show that crop yields correlate with certain climate indicators during successive phenological stages. A subset of these indicators are predictable up to 2–lead months before planting. This opens the possibility to skillfully predict crop yields, without the need to run crop models in ensemble mode. Long forecast lead-times allow anticipatory planning of interventions targeted at each growth stage.

4.1 Introduction

It is possible to predict crop performance ahead of time using climate data as input and a broad range of crop simulation models. This allows assessment of climate and other factors' (Guo et al., 2010) affect on crop production from seasonal, decadal, to climate change timescales (Hui et al., 2013). In many crop models, secondary influences such as soil nutrition, occurrence of weeds, pests and diseases are not implemented. Most research is biased towards direct effects of climate (Guo et al., 2010). Common climate input into crop simulation models are temperature, precipitation, relative humidity, wind speed and incoming solar radiation. Temperature is used to simulate the rate of crop development and other growth processes such as leaf expansion, photosynthesis and respiration. Precipitation, relative humidity, wind speed and solar radiation are used to calculate daily crop water demand. Considering the huge data requirements for crop modelling, forcing by ensemble climate forecasts requires high technical skills. However, this may be mitigated by use of climate indices (i.e. indicators in this study).

Patterns of seasonal climate and their impacts are important for understanding vulnerability and adaptation of regional agricultural production. Robust relationships have been established between regional atmospheric circulations, surface climate and crop productivity (Tao et al., 2008), mainly using the seasonal climate forecasts. Rain fed crop production practiced in the GHA is grossly influenced by meteorological variables (Zhao et al., 2015) and the climate-yield relationships are quite complex. Water deficit is a significant factor in crop production worldwide, but precipitation affects yields non-linearly, implying that good total seasonal rainfall during a crop growing period may not result in good yields (Bannayan et al., 2010; Lobell and Field, 2007). However, quantifying crop vulnerability to water stress is not defined in a standardized and spatially comparable manner (Kamali et al., 2018) and vulnerability differs from one region to another. Linking crop failure to drought in sub–Saharan Africa, Kamali et al. (2018) found geographical dependency of vulnerability to climate stresses.

Vulnerability to water stress (drought) was dominant in South African and Sahelian countries, and temperature stress vulnerability in the Central African countries. For better understanding of spatial differences in maize yield response to climate in the Greater Horn of Africa, we investigate the influence of well-defined climate indicators in two regions of contrasting seasons and climate.

Maize is extremely sensitive to water stress during the critical periods for example from flowering to grain filling (Slingo et al., 2005) mainly due to evapotranspiration and high physiological sensitivity when determining the main yield components such as number of ears per plant and kernels (Omoyo et al., 2015). In North America, higher precipitation is found to have partly contributed to corn yield increases (Nadler and Bullock, 2011). Temperature, drought, wet conditions, and precipitation have been found to have detrimental effects on maize yields in the northern spring zone and Yellow-Huai valley summer maize zones in rural china (Ma and Maystadt, 2017). In southern Africa, within season distribution can affect yields in low rainfall seasons but has little effect in average and above rainfall seasons. Thus, within season rainfall distribution is a critical crop-yield determinant because crop needs water at each growth stage, especially in areas of highly variable and low rainfal (Duffy and Masere, 2015). Also in the GHA it is desirable to assess the influence of rainfall indicators for successive development stages on maize yields.

Temperature effects on maize production are well documented; warmer temperatures accelerate the rate of crop development until an optimum temperature above which development rate reduces, shortening the duration of crop growth and thus reducing yields (Hui et al., 2013). Cooler than normal temperatures slow plant growth rate. Globally, extreme daytime temperatures have been evidenced to have a large negative effect on crop yields (IPCC, 2014). Extreme heats have been found detrimental to crop production in USA (Fisher et al., 2012; Hatfield et al., 2011) and especially when they coincide with critical growth stages. When analyzing the sensitivity of United States maize yield to extreme temperatures, Butler and Huybers (2013) found high temperature sensitivity during silking and grain filling but with major regional variations. In a controlled environment, Hatfield and Prueger (2015) found that warm temperatures increase the rate of phenological development but has no effect on vegetative biomass compared to normal temperatures and that greater impacts of warmer temperatures is in the productive stage as compared to vegetative stage because of a higher vegetative stage optimum temperature. Porter and Semenov (2005) also found damage to crops from increased development stage temperatures. Minimum temperatures affect night time respiration rates and can potentially affect biomass accumulation for example, high minimum temperature increase the rate of senescence and decrease ability of plants to efficiently produce grain (Hatfield and Prueger, 2015). These studies emphasize the effect of climate evolution during maize growth stages.

Climate variables individually and collectively affect crop growth processes contrastingly, often non-linearly at each stage of development. The separate influences aggregate to determine the variability in final yields. Extents of influence of the climate variables vary both temporally and spatially. Thus, understanding the predictability of phenological stages and types of meteorological stresses that affect maize yield variability in each stage may interest many stakeholders in agriculture. In the GHA, predictability of absolute climate forecast variables are highlighted in Ogutu et al. (2017). The ambition of this study is to examine: 1. Rainfall and temperature indicators of significant influence to maize yield variability. 2. Predictability of these indicators, as a function of seasonal climate forecast lead times. 3. Uncertainty in predictability or model simulation of indicators identified in (1.) as a function of forecast lead time. All these are assessed considering the main phenological stages of maize development.

4.2 Methodology

We focus on two sub-national regions of the Greater Horn of Africa (North-Gonder (in Ethiopia) and Bungoma (in Kenya) shown in Figure-4.1 selected based on skillful climate forecasts (Ogutu et al. (2017) and good WLY maize yield predictions (Ogutu et al., 2018). North-Gonder lies between latitudes 11.75N –13.75N and longitudes 32.25E-38.75E, with surface elevation ranging from 527-4000 metres above mean sea level (masl). Rainfall season is in June-September, with mean seasonal rainfall ranging from 470-950mm, average maximum temperature range of 25-36°C and average minimum temperature range of 12-22°C. In the modelling study, seven maize varieties differing in their TSUMs (see Table 5.1) are grown in North-Gonder.

Bungoma spans longitudes 34.3E-35.1E and latitudes 0.43N-1.15N, with elevation of 1249-4000 masl. Rainfall season in Bungoma is in the month of March-May, and October-November with mean mean seasonal rainfall of approximately 700-800mm in MAM; mean maximum temperature range of 25-27°C and minimum temperature range of 12-14°C. We consider only the first cropping season (MAM). Variability in Bungoma's climate is less than the variability in North-Gonder. Only two maize varieties are used in Bungoma, with a single variety with TSUM1=670 planted in over 90% of the area. We focus on this variety and April planting date. We perform analyses based separately on the entire maize growing season, vegetative stage, ±10 days around anthesis, and the reproductive stages.

4.2.1 Climate distribution indicators

We adopt rainfall indicators that describe the temporal distribution of rainfall by the evenness parameter (E_R), the unevenness parameter (U_R) and the shape properties (Ad) of a cumulative rainfall curve as detailed in Monti and Venturi (2007) in addition to rainfall rate (R_R). E_R describes even distribution of rainfall during the growing season, U_R describes uneven independent rain events and Ad enables us to represent either early or late rainfall. In addition we use the total rainfall (R_R) over any period. The indicators E_R, U_R and Ad are used as defined in Monti and Venturi (2007). Similar distribution measures have been used in Duffy and Masere (2015).

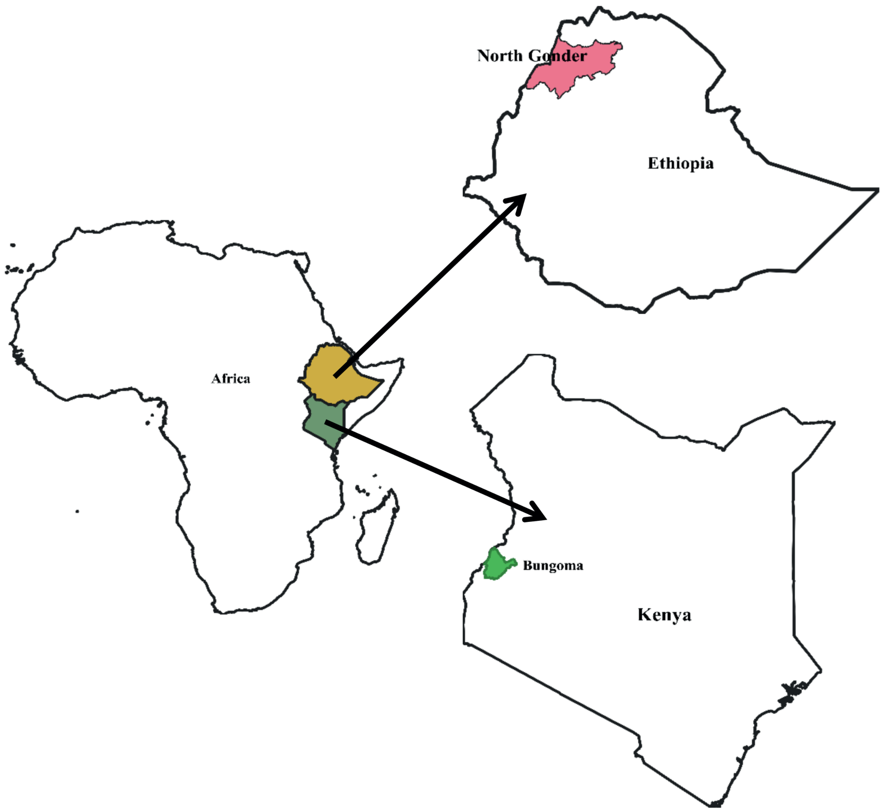


Figure 4.1: Case study areas (North Gonder in Ethiopia and Bungoma in Kenya) shown in relation to Kenya, Ethiopia and African continent

Maize requires a specific amount of heat to transit from one development stage to another. As indices, we use the daily maximum temperature (TX), daily minimum temperature (TN), daily temperature (TG), Growing Degree Days (GDD) and Killing Degree Days (KDD). Heat accumulation during crop growth is assessed using Growing Degree Days (GDD); it estimates the beneficial effects of temperature and can be used to predict plant development stages. The KDD represent the negative effects of high temperatures, quantifying the effects of temperatures to reduce yields. In this study, we use the GDD and KDD versions as defined in Butler and Huybers (2015), considering both the lower threshold (base temperature) and the maximum temperature for maize growth.

4.2.2 Climate and yield data

To assess the relationship between yield variability and weather/climate characteristics, we derive reference climate indicators from the 0.5°× 0.5° gridded Water and Global Change (WATCH) forcing data (WFDEI) for the period 1981-2010. The WFDEI is based on monthly ERA-Interim reanalysis bias corrected using gridded observed climate data from the University of East Anglia's Climate research Unit (Harris et al., 2014). Preparation of WFDEI is detailed in Harding et al. (2011); Weedon et al. (2010, 2011, 2014). Forecast climate indicators are derived from the European Centre for Medium Range Weather Forecasts (ECMWF) system-4 seasonal climate ensemble prediction system (*S4*) reforecasts for the period 1981-2010 described in Molteni et al. (2011), but bias corrected against the WFDEI. The seasonal climate reforecasts is an ensemble of 15-members (individual forecasts), each being produced from a different initial conditions. Details of *S4* and generation of ensembles is found in Molteni et al. (2011).

We generate water-limited reference maize yields from WOFOST driven by WFDEI. In relation to crop physiology, the WOFOST enables a specification of local maize varieties. Originally developed for farm scale use, the model has been adopted for regional simulation with Ogutu et al. (2018) providing details related to the model inputs, setup, and validation against nationally available maize yield statistics for East African countries. WOFOST is setup to simulate water limited yields at a horizontal grid resolution of 0.1°× 0.1° resolved by superimposing Food and Agriculture Organization (FAO) land use data, topography, crop calendar, soil data and climate data. Detailed description of WOFOST model is available in Boogaard et al. (2013); Van Diepen et al. (1989); Keulen and Wolf (1986); Rotter (2014); Supit et al. (2012, 2010). In this study, both yields and climate data are aggregated to sub-national boundaries representative of administrative units (NUTS) for which agricultural yield statistics are collected using the equation:

$$X_{aggregated} = \frac{\sum_{i=1}^{k} X_i \times Area_{cultivated,i}}{\sum_{i=1}^{k} Area_{cultivated,i}} \tag{4.1}$$

Where i = each grid cell within NUTS-2 boundary, k = total number of grid cells, X = simulated yields ($\mathrm{Kg\,ha^{-1}}$) by WOFOST, and $\mathrm{Area_{cultivated,i}}$ = cultivated area. The latter is based on observed production statistics. NUTS concept is explained in Resop et al., (2012) and Supit and Van Der Goot, (1999) i.e. NUTS-0 corresponds to the national boundaries, which are in addition divided into NUTS-1 regions and farther to NUTS-2 (sub-region).

4.2.3 Forecast verification metrics

Three forms of verification metrics are chosen, i.e. Metrics that work on the full ensemble members to assess the spread of the forecasts (spread Vs. Skill ratio) and to assess consistency and reliability (the rank histogram); metrics that work on probabilistic forecasts such as the relative operating curve skill score (ROCSS) and the generalized discrimination score (GDS); and Error metrics that work with the mean (or median) of the ensembles and [i.e. mean error (bias), percentage bias (pbias), and the root mean squared error (RMSE)]. The metrics are determined as implemented in Bhend et al. (2016).

The spread-error ratio (SprErr)>1 implies an over dispersion (under confidence) i.e. observed variability is higher than the forecasted variability. SprErr <1 1implies under dispersion (over confidence) i.e. the observed variability is lower than the expected variability (under dispersion/over confidence), and SprErr = 1 implies that there is no discrepancy between the observed and forecasted variability (equi–dispersion). We use the metric to measure uncertainty.

The ROCSS, derived from the area under the Relative Operating Curve (AROC) measures skill of tercile forecasts i.e. above normal (upper tercile), below normal (lower tercile) and near normal (middle tercile) forecasts. Briefly, ROCSS = 2AROC -1, where $-1.0 \leq \text{ROCSS} \leq 1.0$. ROCSS = 1.0 indicates a perfect forecast system; ROCSS< 0 indicates perfectly useless forecast system, and ROCSS = 0 indicates no skill. Detailed descriptions of the ROCSS is available in Barnston et al. (2010); Diro et al. (2012); Hamill and Juras (2006)arnston et al. (2010); Diro et al. (2012); Hamill and Juras (2006); Mason and Graham (1999) and Appendix-B.1.

The generalized discrimination score (GDS) measures how well a forecast can distinguish between two observations; it measures forecast skill. For a skillful forecast, the proportion of correctly picked observations will exceed 50% (i.e. GDS >50%); a perfect score is 100%. Details and formulations of the GDS is found in Ebert et al. (2013); Mason and Weigel (2009); Weigel and Mason (2011); Bhend et al. (2016).

Forecast Bias (bias), Percentage Bias (pbias), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and correlation (r) assess the difference between the observed and ensemble mean or median. The metrics are calculated for each ensemble member individually (15-ensembles) before aggregating the mean over all ensembles. For each ensemble member, the metrics are determined as illustrated in Willmott (1981, 1982).

Like the spread error ratio, the rank histogram is useful for determining the reliability of ensemble forecasts and diagnosing errors in its mean and spread. For a reliable n–member ensemble forecast, if the forecast and observation are pooled into a vector and ranked from lowest to highest then the observation is equally likely to occur in each of the n+1 possible ranks resulting in a flat histogram. Consistent bias in the ensemble will show a sloped rank histogram while a lack of variability in the ensemble will show a U-shaped or concave population of ranks (Elmore, 2005; Hamill, 2001). Statistical tests are used to assess flatness or otherwise of the ranks and to identify the causes of deviation. We use the chi-square test (X^2) to assess deviation of the resulting rank histogram from flatness and its decomposition explained in Jolliffe and Primo (2008) to identify specific alternatives to flatness such as bias or under/over dispersion. A detailed description and discussion of the rank histogram is available in Elmore (2005); Hamill (2001); Jolliffe and Primo (2008).

4.3 Results

4.3.1 Ethiopia

4.3.1.1 Climate indicators that significantly correlate with maize yield

Figure 4.2; rows 1 and 2 show rainfall and temperature indicators that explain maize variability in North-Gonder (northern Ethiopia). Considering the entire growing season, the average rainfall rate (R_R), evenness parameter (E_R), average season minimum temperature (TN), daily temperature (TG), explain yield variability. Growing season R_R, E_R, TG, and KDD respectively explain 15%, 16%, 26%, and 8% variability in yields. This illustrates the importance of both rainfall and temperature indicators to maize yields.

In the vegetative stage, R_R, E_R, timing Ad and TX show significant correlations with yields and respectively explain 21%, 20%, 20%, and 15% variability in yields. More rainfall indicators during vegetative stage explain yield variability than does temperature indicators.

Around anthesis, Ad, TN and TG have significant correlation to yield and respectively explain 21%, 19%, and 17% of yield variability. In the reproductive stage average maximum temperature TX, TN, TG, and GDD respectively explain 17%, 11%, 23%, and 24% of maize yield variability. In anthesis and reproductive stage, it is noted that temperature explain maize yield variability as opposed to rainfall indicators.

4.3.1.2 Predictability of indicators

Figure 4.2 and Figure 4.3; row 5 show predictability of above-normal and below-normal climate indicators for northern Ethiopia using ROCSS. We use significance of the ROCSS (shown by predictability threshold line) to identify predictable characteristics.

Growing season R_R, E_R, TN and TG are identified in section 4.3.1.1 to influence yields. Above-normal forecasts of the indicators are predictable with significant skill in all the forecast lead-times. Predictability of below normal forecasts of these indicators vary. For example, below normal R_R forecasts is predictable at lead-0, and -2 while E_R is predictable at lead-1 and lead-2 only. Below normal forecasts of Daily temperature (TG) are predictable in the three lead-times. Growing season KDD forecasts are predictable with significant skill. Even though not identified to influence yields, its predictability provides information that may be useful in developing mitigation options against effects of high temperatures. Other verification metrics are shown in Table D.1. All growing season forecast indicators

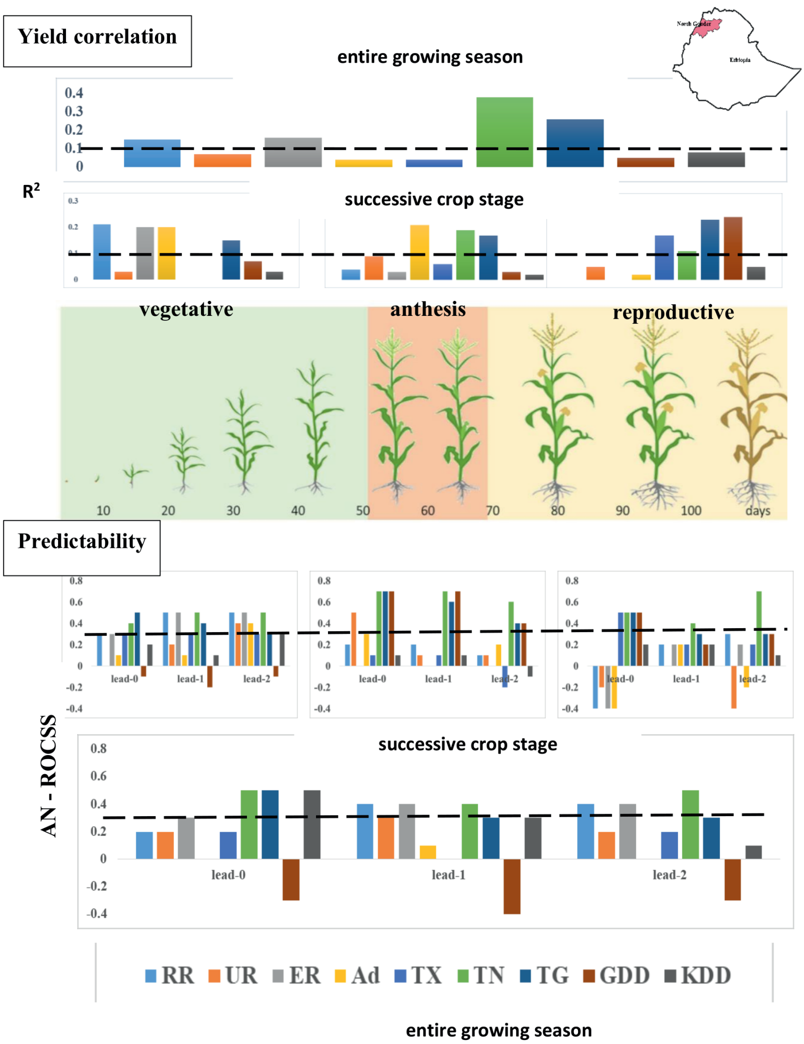


Figure 4.2: Maize yield variability as explained by various climate indicators (top 2 rows of bars) and the predictability of the latter (bottom 2 rows of bars) for both the entire June –July-August growing season and the three crop development stages within that season (schematized in middle graph) for North Gonder, Ethiopia (map insets). The wide bars at far top and far bottom present the respective statistics for the entire growing season, the narrow bars on the second and before last rows reflect the three consecutive development stages, from left to right Vegetative, Flowering (Anthesis) and Reproductive stage, respectively. The metric used to quantify explanatory power of the indicators on yield is here the Coefficient of Determination R^2 (other explanatory metrics are given in supplementary tables D.1 to D.4 The metric used to quantify predictability is the ROCSS shown in this figure for the upper tercile, or above normal (AN) conditions. Significance levels for correlation (r^2>0.1) and skill (ROCSS>0.33) are shown with black dashed lines with apparently missing bars implying negative correlation coefficient and or zero ROCSS respectively.

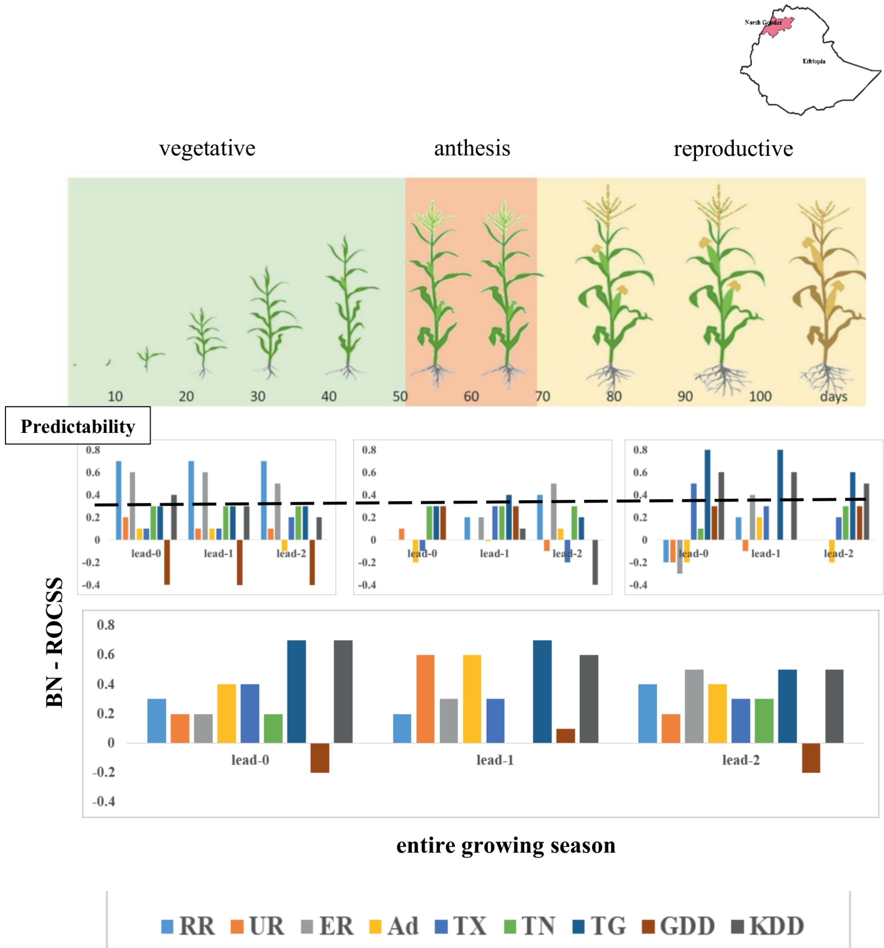


Figure 4.3: Same as figure-4.2, for Below Normal Indices; correlations (top two rows) are not shown are not shown (identical to figure 4.2)

including those not found to influence yields have good forecast discrimination shown by GDS>50% (Table D.1)in all lead-times except GDD which has poor discrimination (i.e. GDS = 45%, 46% and 40% in lead-0, lead-1 and lead-2 respectively). Errors in forecasts quantified by pbias, MAE, and RMSE in Table D.1 are low in all lead-times with pbias$\leq$ 10%. This is also true for all growing season climate indicators displaying poor terciles forecasts. Temperature characteristics show higher correlations with the observed at all lead times except GDD. During the growing season in North-Gonder, climate indicators that influence yields have variability almost equal to that of the observed (equi-dispersion) as shown by SprErr in Table D.1. Rainfall indicators especially have higher SprErr in all lead-teams indicating better spread than that of temperature indicators. GDD and KDD are characterized by under-dispersion and over-dispersion respectively in all lead-times. Good dispersion in rainfall indicator forecasts are also shown by uniform rank histograms at lead-0 and lead-1 (Table D.9). Lower R_R and E_R, (SprErr ≈ 0.7) at lead-2 translates to non-uniformity of rank histogram as a result of either biases or lack of variability or both. Forecasts of growing season temperature indicators are uncertain while all forecasts of rainfall indicators show equi-dispersion and uniformity of histograms irrespective of lead-times (shown in Table D.1 and Table D.9) implying certainty in all the three lead-times.

In the vegetative stage (row-4 of Figure 4.2 and Figure 4.3), both AN and BN forecast of R_R, E_R, and TG are predictable with significant skill in all the three lead-times. These indicators are identified to explain yields together with Ad. But Ad is unpredictable in all lead-times and all terciles. It shows the difficulty in predicting the time of occurrence of rainfall (Ad) during vegetative stage. TN, which has no influence on yields is predictable in all terciles with significant skill suggesting its importance for other crops. For example, crops that are sensitive to frost like conditions such as tea. Forecasts of vegetative stage climate indicators are in Table D.2. All indicators have good discrimination with GDS ranging from 52% to 71% in all lead-times except GDD. GDD also show no correlation with the observed nor is it predictable. There exist acceptable biases in forecasts of all climate indicators that influence yields as seen in the error statistics in (Table D.2). Ad, and KDD especially at lead-1 and lead-2 show high biases rendering them unsuitable for yield estimation. Results show that vegetative rainfall indicators are more predictable compared to temperature characteristics. The predictability and significant correlations with the observed characteristics imply the possible use of such to forecast yields when actual forecasts are unavailable. Forecasts of both rainfall and temperature indicators during vegetative stage show reasonable spread i.e. near equi-dispersion (Table D.2). GDD is grossly under-dispersed in all lead-times while KDD forecast is over-dispersed at lead-2. Rainfall characteristics show uniformity of rank histogram (Table D.9) except lead-2 E_R forecasts that deviate from uniformity due to bias. All temperature indicators deviate from uniformity of histogram in all lead-times due to either bias or lack of variability or a combination of both. The findings imply a higher uncertainty in forecasts of temperature indicators forecast compared to rainfall.

Of the anthesis period indicators that influence yields, only TN and TG are predictable with significant skill in all terciles and lead times. Though anthesis period GDD does not influence maize yields, it is predictable with significant skill (shown by ROCSS). Predictability of GDD is useful as it makes it possible to forecast ripening stage or disease and pest outbreaks hence implications for farm management practices. Temperature indices that influence yields, plus all rainfall indicators exhibit good forecast discrimination in all the three lead-times. However, discrimination of rainfall indicators are lower (see GDS in Table D.3). KDD forecasts show poor discrimination of GDS = 48%, 50% and 42% in lead-0, lead-1 and lead-2 respectively. Apart from TN and TG showing significant correlations with the observed in all lead-times, the rest of the characteristics have almost nil correlations with the observed. But even with the low correlations, simulation errors as shown by the bias, pbias, MAE and RMSE are within acceptable limits in all lead-times except lead-2 KDD with pbias of -42% (pbias in Table D.3). Considering predictability and error statistics, most of the climate indicators during anthesis can be useful in yield estimation. Forecasts of TX (lead-0 and lead-1), and GDD (lead-0 to -2) exhibit under-dispersion (shown by SprErr). U_R, and E_R are under-dispersed at lead-0. Rainfall characteristics forecasts have reasonable dispersion at longer lead times. Under-dispersion of GDD forecasts in all lead-times implies nil certainty. The rank histogram statistics (Table D.9) show that GDD forecasts deviate from uniformity due to biases in simulation in all lead-times. Generally, GDD forecasts around anthesis have higher uncertainty in prediction compared to all other characteristics irrespective of the forecast lead-time.

In the reproductive stage, TX, TN, TG and GDD are found to influence yields. Of these above-normal forecast of TN is predictable across the three lead-times while that of TX forecast is predictable at lead-0 only. Both above-normal and below-normal forecasts of TG and GDD have prediction skill. Rainfall indicators during the reproductive stage are not predictable. TX, TN, TG and GDD indicators identified to explain yield variability all have good forecast discrimination (GDS>50%) in the three lead-times (see Table D.4). Lead-0 rainfall characteristics forecasts have no discrimination (i.e. GDS < 40%) but this improves in longer lead-times. Rainfall characteristics also show very low correlations at lead-0 but this improves at lead-1 where there exist significant correlations. Temperature characteristic forecasts have significant correlations with the observed at lead-0 but this reduces in longer lead-times. Temperature indicators are dominated by negative biases in all lead-times. The errors are acceptable except lead-1 Ad with pbias of 29%. Rainfall characteristics have higher simulation errors and less predictable compared to temperature indicators.

All indicators in the reproductive stage have reasonable spread as shown by the SprErr in Table D.4 with the exception of GDD (under-dispersion in lead-1) and KDD (over-dispersion in lead-1 and lead-2). All forecasts of rainfall and temperature indicators show uniformity in rank histogram except TX (lead-0 and lead-1), GDD (lead-1and lead-2) and TG (lead-2). Lead-1 GDD rank histogram deviation is due to both a lack of variability and bias in the forecasts while the rest of deviations from uniformity are due to bias. Generally, forecasts of rainfall indicators in North-Gonder during consequent phenological stages and in the entire crop cycle may be predicted with higher certainty at lead-0 and lead-1. Temperature characteristics show certainty except GDD and KDD though this varies with lead-time and stage of growth.

4.3.2 Bungoma

4.3.2.1 Climate indicators that significantly correlate with maize yield

In the equatorial region of East Africa (i.e. Bungoma; Kenya), rainfall indicators do not have significant correlations with yields irrespective of the stage of growth (Figure 4.4a snd 4.5b, rows 1 and 2). Temperature indicators in the entire growing season, vegetative stage, and around anthesis explain the variability in yields but marginally in the reproductive stage. Growing season TX, TN, TG and KDD account for 39%, 28%, 41% and 9% of the variability in yields. Temperatures during the vegetative stage seem to account for a higher percentage in variability i.e. TX, TN and TG account for 58% of variability in yield while KDD accounts for 23%.

Around anthesis, TX, TN, TG, GDD, and KDD individually account for 32%, 37%, 36% and 20% of yield variability. In the reproductive stage however, TX, TN, and TG marginally explain yield variability. Yield variability cannot be explained by rainfall indices because there could be enough water or rainfall is regular. Sensitivity to temperature as opposed to soil moisture (in extension rainfall) has also been found in Iizumi et al. (2013).

4.3.2.2 Predictability

The ROCSS shown in Figure 4.4 and Figure 4.5 illustrate AN and BN predictability of climate indices as a function of forecast lead-time respectively for the entire growing season and maize growth stages in Bungoma. In the growing season and reproductive stage, we show predictability for lead time-0 and lead-1 because forecasts beyond lead-1 do not span the entire duration of maize growth. Growing season AN and BN rainfall indicators lack predictability but TX, TN and TG forecasts are predictable in lead-times zero and one. Though GDD and KDD do not explain yields, their forecasts are predictable and may be useful for farm management practices.

Forecasts of growing season temperature indicators that explain yields (i.e. TX,TN, TG) and indeed all forecast of temperature indicators exhibit good discrimination, low biases in all lead times considered and good correlations with the observed. GDD shows poor forecast discrimination, poor correlation with the observed but with low forecast errors in the three lead-times. Forecasts of rainfall indicators generally exhibit poor discrimination, poor correlations, and higher errors in all lead-times as evidenced by bias, pbias, MAE, and RMSE (Table D.5).

Yield explaining temperature indicators show almost equi-dispersion in all lead-times except TN with SprErr = 0.5 in both lead-0 and lead-1 respectively (see Table D.5). Maximum temperature (TX) SprErr degrades from equi-dispersion (SprErr = 1) in lead-0 to under-dispersion (SprErr = 0.3) in lead-1. Forecasts of rainfall indicators are almost equi-dispersed in both lead times. Ad, TN, and GDD rank histograms deviate from uniformity in lead-0 and lead-1 due to either bias or dispersion as shown in Table D.10. The three indicators are already seen to have SprErr$\leq$0.6. Generally, it is observed that in many instances. SprErr $\neq$1 results into deviations from uniformity of histograms. Of the growing season climate indicators that influence yields, TX and TG are forecasted with certainty as are rainfall characteristics except Ad.

In the vegetative stage, TX, TN, TG and KDD explain variability in maize yields and are predictable in lead times-0, 1, and 2. GDD though not found to explain yield variability, is predictable as are some rainfall indicators. For example, above-normal and below-normal forecasts of E_R are predictable in the three lead-times, R_R in lead-1 and lead-2, while Ad is only predictable at lead-0. Vegetative stage predictability assessment suggests that variability in temperature indicators is feasibly more useful in predicting yields than rainfall indicators.

Forecasts of yield explaining temperature indicators (i.e. TX, TN, TG, KDD) all exhibit good forecast discrimination (GDS>50%), low forecast errors and good correlations to the observed in all lead-times but not GDD. Negative biases dominated forecasts of temperature indicators (Table D.6) even though they explain yields. Unpredictability of rainfall indicators forecasts is exhibited but errors are low as shown by bias, pbias, MAE, RMSE. Vegetative period Ad is not predictable in all lead-times. In the vegetative stage, all indicator forecasts have variability that is almost equal to the observed as shown by SprErr in Table D.6 in all lead times. KDD forecast show equi-dispersion at lead-0 but over-dispersion at longer lead-times though the rank histogram does not deviate from uniformity (Table D.10). TN rank histogram deviate from uniformity due to both biases and lack of variability in all lead times. All indicators show forecast certainty except TN and KDD.

Around anthesis, only TN and TG forecasts are predictable in both categories (i.e. AN and BN) in all the three lead-times (Figure 4.4 and 4.5; row-4; column-2). TX, GDD, and KDD though found to explain yields are not predictable except below normal forecasts of TX. Some rainfall indicators show insignificant below-normal predictability. All temperature indicators have good forecast discrimination and acceptable biases in all lead-times except KDD (Table D.7). Correlation between forecasts and observations is lacking in all lead-times except TN that with significant correlation of r = 0.4 in all lead times. Ad and KDD have high errors rendering them unusable (i.e. Ad pbias = -98%, -88%, -202%; KDD pbias = 149%, 169% and 157% in lead-0, lead-1 and lead-2 respectively). Lead-0 R_R and E_R also possess higher biases (>25%). Only KDD ensemble forecasts have a variability that is higher than the observed (i.e. over-dispersion) in all lead-times implying uncertainty. The rest of the forecasted indicators show similar variability to the observed in all lead-times. But considering the rank histograms, only forecasted lead-0 R_R show a deviation from uniformity of the rank histogram at lead-0 (Table D.10) due to bias even though the variability is equal to that of the observed. Even with low uncertainty, lead-0, lead-1 and lead-2 biases in R_R, E_R, Ad and KDD forecasts have high forecast errors (pbias in Table D.7) thus rendering them unusable. With thse findings, we propose that a single metrics is not enough to assess all attributes of a forecast.

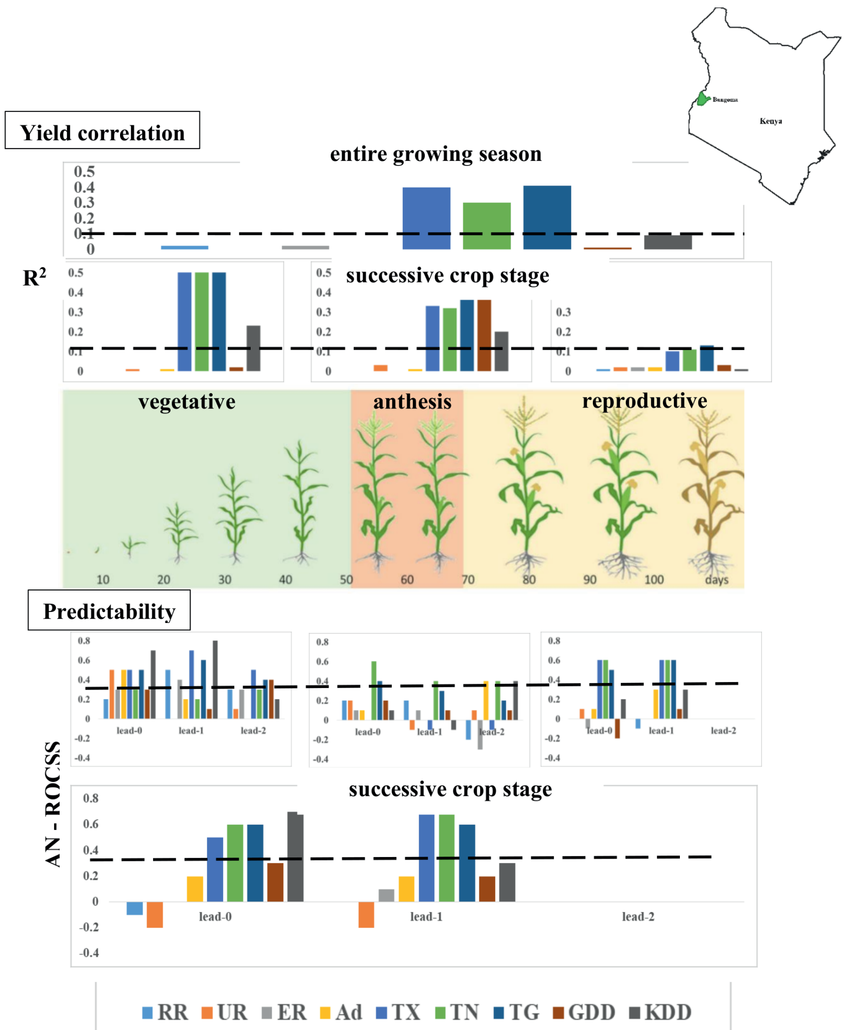


Figure 4.4: Maize yield variability as explained by various climate indicators (top 2 rows of bars) and the predictability of the latter (bottom 2 rows of bars) for both the entire June –July-August growing season and the three crop development stages within that season (schematized in middle graph) for North Gonder, Ethiopia (map insets). The wide bars at far top and far bottom present the respective statistics for the entire growing season, the narrow bars on the second and before last rows reflect the three consecutive development stages, from left to right Vegetative, Flowering (Anthesis) and Reproductive stage, respectively. The metric used to quantify explanatory power of the indicators on yield is here the Coefficient of Determination R^2 (other explanatory metrics are given in supplementary tables D.5 to D.8. The metric used to quantify predictability is the ROCSS shown in this figure for the above normal (AN) conditions. Significance levels for correlation (r^2>0.1) and skill (ROCSS>0.33) are shown with black dashed lines with apparently missing bars implying negative correlation coefficient and or zero ROCSS respectively.

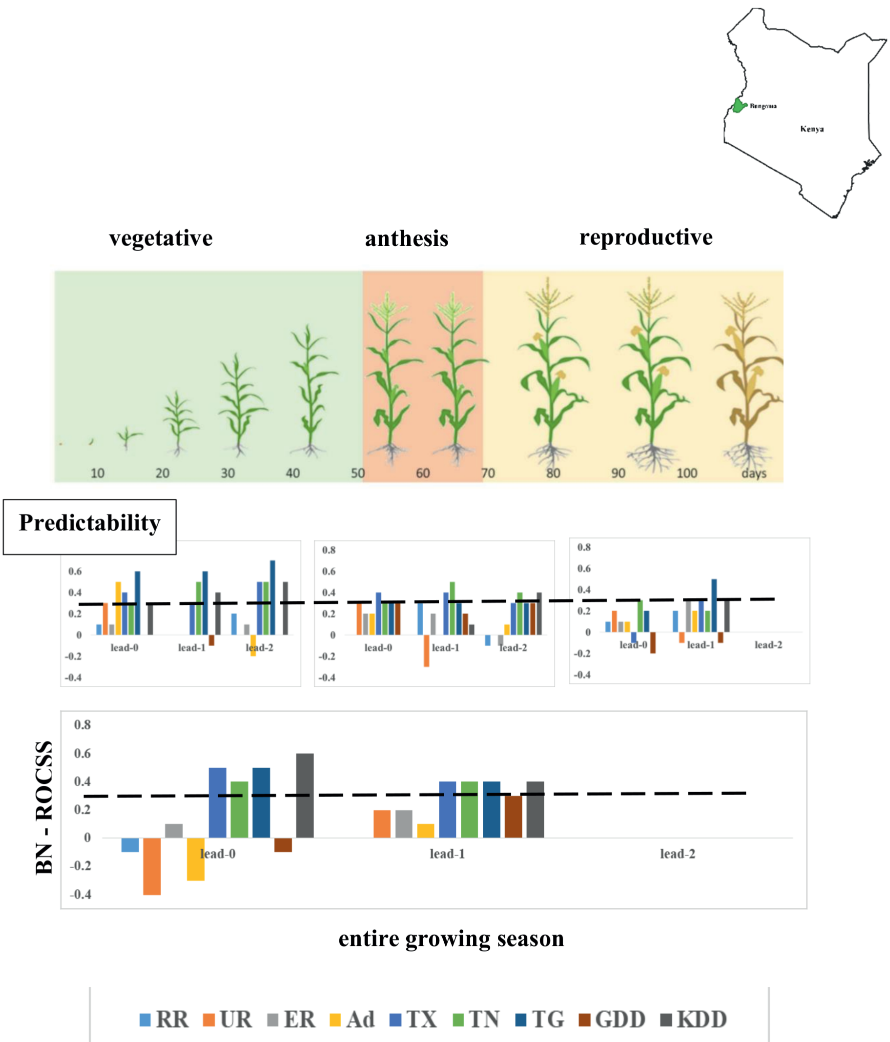


Figure 4.5: Same as Figure-4.4 for Below Normal Indices; correlations (top two rows) are not shown (identical to 4.4)

Only temperature indicators are predictable in the reproductive stage i.e. above-normal TX, TN and TG are predictable (with significance) at lead-0 and lead-1 while the BN predictability do not show significant ROCSS. Important indicators (i.e. TX, TN and TG show good forecast discrimination, low forecast errors (shown by bias, pbias, MAE, RMSE) at lead-0 and lead-1 (Table D.8). All reproductive stage rainfall indicators exhibit poor forecast discrimination and no predictability. As with the other growth stages in Bungoma, forecasts of rainfall indicators have higher simulation errors compared to temperature. In the reproductive stage, TN forecasts are under-dispersed in lead times zero and one but the rank histogram does not deviate from uniformity. GDD and KDD are over-dispersed but the rank histogram does not deviate from uniformity. Even though forecasts of rainfall indicators exhibit under-dispersion, the rank histograms do not deviate from uniformity. Only TX and TG are equi-dispersed and rank histograms do not deviate from uniformity in lead times-zero and –one implying only TX and TG are forecasted with certainty. The findings for reproductive stage also highlights two challenges, one; one metric is not enough to measure all attributes of a forecast, and two; more than metrics for a single attribute makes it difficult to conclude on the performance of an attribute. A compromise should therefore be allowed depending on the attribute of interest.

4.4 Discussion

4.4.1 Methodology

Agriculture in East Africa is majorly rain fed and therefore the importance of good rainfall amount cannot be over-emphasized. Rainfall distribution during a crops life may play a greater role in crop yields than the total rainfall because rather than rainfall amounts, available soil water determines maize growth and yields as discussed in Asseng et al. (2001); Cakir (2004); Duffy and Masere (2015); Huzsvay and Nagy (2005); Monti and Venturi (2007). Few studies have examined the influence of rainfall indicators on yields in East Africa. This study both supports and makes use of these rainfall indices to quantify the timeliness and rainfall distribution over the course of maize growth. It allows a possible quantification of their influences on maize yields. We show clearly that both rainfall rate ($\mathrm{R_R}$ and rainfall distribution ($\mathrm{E_R}$ and Ad more so than $\mathrm{U_R}$) are important in areas where crop production is more water than temperature limited, such as Northern Ethiopia. The distribution indicators are obviously correlated. $\mathrm{R_R}$ and $\mathrm{E_R}$ should covary strongly, so only one could be used.

When we consider their relation with yields, correlations of $\mathrm{R_R}$ and $\mathrm{E_R}$ with yield are generally very similar, as are those of $\mathrm{U_R}$ and Ad even though the two do not covary. Also when we consider their predictability, their similarities are obvious, $\mathrm{R_R}$ and $\mathrm{E_R}$ have practically equal skill across all metrics, as do $\mathrm{U_R}$ and Ad. So, yes, when building a climate service, providing $\mathrm{R_R}$ or $\mathrm{E_R}$ and $\mathrm{U_R}$ or Ad would suffice. Perhaps, some measure of extreme rainfall could have been added covering the hazardous effects of extreme precipitation events on yield.

Temperature is known to influence maize yields especially in terms of the number of days in which daily temperature is within the lower and upper limits of maize growth or above the upper threshold represented by GDD and KDD. Apart from the actual temperatures, the temperature indices represent temperature influences on plant physiological processes which in turn determine biomass assimilation and allocation without details of bio-physiological processes. Also here the indicators used must be co-varying to some extent, though perhaps less directly than the ones for rainfall. TX and KDD both represent heat stress, and TG accumulates in GDD when above the base temperature, but TN or TX relation to GDD is not so straightforward. Looking at the results, both at their correlation with yield, or at their predictability, their co-variance is apparently lower than perhaps expected and we cannot simply discard one in favor of another. However, GDD and KDD are because of their cumulative nature, very sensitive to biases in relation to the thresholds used in their calculation. That makes data with relatively low spatial resolution notoriously problematic in regions with intense topography (applicable in North Gonder more so than in Bungoma). Assessing influences of rainfall and temperature indicators on yields individually at various phenological stages may enable design of mitigating inventions because of negative influences. For example, use of drought resilient cultivars when depressed rainfall is forecasted or quick maturing varieties in hotter areas or shifting to a more resilient crop if climate conditions are dire.

Forecasts with long forecast lead times before planting allow design of farm management practices and can be adjusted with reduction in lead time. Ensemble forecast enables quantification of uncertainty in the forecasts as it is already known that seasonal forecasts are imperfect. We have used a number of verification metrics to sample aspects of forecast attributes since not all may be good. For example, the terciles of a forecast may be predictable (assessed by ROCSS) yet possess errors (biases) and may/may not be of similar variability with the observed.

4.4.1.1 Yield explaining climate indicators

We have identified rainfall and temperature indicators that explain maize yield variability over two sub-national regions differing in seasons and season length, i.e. North-Gonder in northern Ethiopia and Bungoma in the equatorial region of Kenya. It has allowed a view of spatial differences on climate-yield relationships. Growing season rainfall and temperature are the main climatological variables influencing maize phenology and yields. The relationship between either temperature and yields or rainfall and yields are generally inhomogeneous in the two locations. While both rainfall and temperature indicators influence yields during a crops' growing season in northern Ethiopia, only temperature indicators influence yields in Bungoma. These relationships between climate variables and maize yields are location specific and cannot be generalized.

Climate-yield relationships may result from other factors such as temperature-rainfall interactions, soil types and hence water holding capacity, temperature-water vapor deficit interactions. These findings are not unique, as similarities have been obtained in both global and regional studies. For example, in studying the impact of climate variation in global crop yield variability, Ray et al. (2015) found that where, and by how much a crop's yield varied due to climate is highly locational and crop specific. Huang et al. (2015) found that total growing season precipitation had a stronger influence on yields in North eastern United States (US) than in the South East. Varying sensitivity to climate variables has also been found in Mongolia (Huang et al., 2017); North East China (Zhao et al., 2015); Sri-Lanka (Karunaratne and Wheeler, 2015). In sub-Saharan Africa, sensitivity to water stress is found in southern Africa and some countries in the Sahel region, while sensitivity to temperature stress is in Central African countries (Kamali et al., 2018).

Yield sensitivity to growing season temperature only, as in Bungoma may be due to uniform and sufficient precipitation and soil moisture variability during the crop season i.e. there exist less spatial and temporal variability as opposed to northern Ethiopia. This may be true as the entire Bungoma almost falls within one grid box of $0.5^{\circ} \times 0.5^{\circ}$ ERA-Interim input climate data as opposed to the larger North-Gonder region with many grid cells that could result in a higher spatial variability.

Both vegetative stage rainfall and temperature are important for maize growth. We however find contrasting yield sensitivities in northern Ethiopia and equatorial Kenya. In northern Ethiopia, even distribution and period of occurrence of rains, the average daily temperature, GDD and KDD during vegetative growth stage influences yield variability.

In equatorial Kenya however, yields show sensitivity to vegetative stage temperature indicators only (i.e. maximum temperature, minimum temperature, mean daily temperature and KDD). Higher daily temperatures during this period result in higher respiration rates and faster growth rate (Via influence on GDD). Higher temperatures above the upper threshold influences KDD, which in turn limits carbon assimilation. Sufficient water availability lowers yield sensitivity to temperature. This may be the reason for lower R^2 between temperatures and yields in North-Gonder.

A difference in sensitivity of yield is evidenced during anthesis. While the time of occurrence of rainfall (Ad), minimum temperature and average daily temperature during anthesis influence yields in northern Ethiopia, only temperature indicators influence yields in Bungoma. Maize has high sensitivity to high temperatures during anthesis and in the reproductive stage i.e productivity is reduced when extreme temperatures and is further exaggerated when there are water deficits (Hatfield and Dold, 2018). This may explain sensitivity to Ad (timing of occurrence of rainfall) amongst the rainfall indicators in northern Ethiopia. Regional variation in sensitivity of yield to temperatures during silking and reproductive stage have been found in the USA (Butler and Huybers, 2013) like in this study and with a higher sensitivity in the reproductive stage (Hatfield and Prueger, 2015).

During anthesis and reproductive stages, warmer temperatures above upper threshold causes a reduction in yields as it accelerates development thus reducing the period required for grain filling and reduces rate of photosynthesis. This statement may explain the sensitivity to temperature characteristics (anti-correlations) as opposed to rainfall during anthesis and development stage in the two regions. In other studies however, water availability during anthesis and grain filling stage have been found highly explanatory of-and- results in high grain yield (Armstrong, 1999; Butts-Wilmsmeyer et al., 2019). Because both rainfall and temperatures are important in all stages of maize growth, we hypothesize that regionally varying sensitivities may be a function of other factors not examined such as soil characteristics and soil moisture availability during each phenological stage. In conclusion, sensitivity of yields to climate characteristics is localized.

4.4.2 Predictability

Generally, R_R (and thus E_R) can be better predicted than E_R (or Ad), although there is skill for the latter sometimes in the vegetative stage (which is effectively closer to initialization of the forecast, i.e. a reduced lead time). Similarly, TN, TG and TX can be predicted well than KDD, while GDD cannot be predicted at all in the two regions. Not all indicators are predictable in all growth stages nor forecast terciles and forecast lead-times.

The results are interesting because higher predictability is expected at shorter lead-times (lead-0 for example) than at longer lead months but this is not the case in our findings. Some indicators are less predictable at lead-0 compared to lead-1 and lead-2. Comparable findings are seen in Ogutu et al. (2017) in which for example, in region-4 where Bungoma falls, MAM lower-tercile rainfall is predictable in lead-times 1-2 but only lead-1 upper-tercile is predictable. And in northern Ethiopia (region-A) only lead-0 lower-tercile rainfall forecast show some skill while the upper-tercile forecasts are not predictable at all. The findings do not negate findings in this study but may offer advantage of longer lead-time forecasts.

We find good predictability of daily temperature at all the three lead-times agreeing with findings in Ogutu et al. (2017). Likewise, minimum and maximum temperatures are more predictable than rainfall indicators. Predictability of indices related to temperature quantities (i.e. GDD and KDD) is poorer but given their collinearity to daily temperature and maximum temperature respectively, predictability of the former is still useful in all the crop development stages. Since we have not identified days with temperatures higher than the upper threshold for maize growth, we may speculate that predictability of growing season temperatures or GDD and KDD may give farmers a chance to prepare for adaptation options early in the season for example by providing irrigation during periods of high temperatures, it noted in Anderson et al. (2015) that sensitivity of maize yields to high temperatures may be modulated by availability of sufficient moisture.

Predictability of indicators during each growing phases offer possibility of short time tailored interventions. In vegetative state, high temperatures accelerate the rate of crop development, one reason being that the number of thermal units mandatory for leaf appearance is invariable (Hatfield and Dold, 2018) while increased rainfall intensity during vegetative stage has been found to decrease yields (Mtongori et al., 2015). The presence of good predictability of above and below-normal vegetative stage rainfall distribution and daily temperatures in lead months zero to two in North-Gonder provide good potential to assess maize yields. Low temperatures decrease biomass and leaf area index because of low photosynthesis and carbon assimilation rates. Predictability of periods of low temperatures before planting offer the possibility to adapt by for example changing plant dates to limit low temperature effects.

In this study, rainfall indicators during anthesis do not influence yields nor are they predictable in both North-Gonder and Bungoma. It should be realized that the anthesis period is much shorter than either vegetative or reproductive stage, which makes it inherently more difficult to predict its weather at long lead times. This does not however invalidate the importance of rainfall indicators during anthesis. Rather, the influence of rainfall attributes may be muted by adequate soil moisture since both water availability during anthesis and extreme temperature episodes affect maize yields (Butts-Wilmsmeyer et al., 2019; Hatfield and Dold, 2018). Anthesis period average minimum temperature and daily temperature are predictable in all the three lead-times in both North-Gonder and Bungoma. It is known that exposure to high temperatures (best represented by KDD in this study) during anthesis reduces kernel number and hence grain yields. KDD is not predictable. Even so, the effects of high temperatures can still be mitigated using the predictability of daily temperatures (TG).

We did not identify significant correlations between reproductive stage rainfall nor temperature indicators with yield variability in Bungoma. The exists however, a significant correlation and predictability of the daily temperature with three months lead-time in North-Gonder in which also TX and GDD have predictability at only lead-0. The finding agrees with other studies showing sensitivity of yields to reproductive stage temperatures. For example, exposure to elevated temperatures during this stage reduces yields and the impact is even higher with water deficits or excess soil moisture (Hatfield and Prueger, 2015). This may still be useful in assessing yields.

4.4.3 Uncertainty in predictions

Seasonal climate forecasts are characterized by uncertainty, in many instances, ensemble forecasting samples the uncertainty space and provides forecast in terms of probabilities. Measures such as ensemble spread, and its relationship to ensemble error (SprErr), the spread of the histogram amongst others also quantify uncertainty. The two measures are interrelated, i.e. non-uniformity of rank histogram is due to bias or due to difference in variances between the observed and the forecasts (or both). It has been noted in this study that SprErr lower or larger than unity results in non-uniformity of the rank histogram thus the two measures may result in discrepancy when used together. Though the need to measure different attributes of a forecast has been advanced, we would propose not to use many measures for a single attribute. An example is forecasts of GDD in North Gonder which shows consistent non-uniformity of histograms (uncertainty) in all growth stages considered though with acceptable spread-error ratios (certainty) in some stages. Some characteristics (e.g. North-Gonder rainfall indicators) are predictable with certainty irrespective of the measure used indicating robustness in the forecasts. Uncertainty in climate characteristics prediction depend on region and on maize growth phase but varies less with forecast lead-time. For example, KDD uncertainty in predictions vary with phenological phase and geographical region. Generally less uncertainty ±10-days around anthesis may be due to short duration in terms of days. Though far from the time of lead-0 forecasts (e.g. > 60days), interventions at this critical phase of maize growth can still affect maize yields.

4.4.4 Implications of the study

Precipitation and temperature are the two critical climate factors but not the only factors that influence maize production. They are most critical in rain fed agriculture in East African region. Assessing the influence of their characteristics in terms of indicators/indices during the maize growing season and in the successive phenological stages provides a better understanding of the role of climate on certain (critical) stages of maize growth.

This would enable development of geographical region and phenological stage targeted interventions to maize production. Understanding of yield response to varying climatic conditions and partial variation in sensitivities is crucial for sustainable maize production since it informs adaptation and increases resilience to maize production especially in the current climatologically changing world. To researchers, identifying suitable climate indicators during maize growth phases that influence yields makes it possible to harvest expected yields depending on weather and adaptation practices during each growth stage. Since it is the first term rainfall indices defined in Monti and Venturi (2007) are used over eastern Africa, (i.e. E_R, E_R, and Ad), the study contributes to scientific body of knowledge. Changes in expected yields are detectable with each growth phase. It also enables the use of simple climate indices to forecast yields, while foregoing the use of complex process based crop models, driven by daily forecast data, in the forecast production chain. For organizations that offer agriculture related climate service in the region such as IGAD Climate Prediction and Applications Centre (ICPAC), the results of this study may enrich their products such as the crop monitor products.

4.5 Conclusion

The general conclusion in this study is that the relationship between climate indicators and maize yields are location specific, and differ with crop growth stage and forecast lead-time. As such for possible use in yield forecasting, useful indicators should be chosen according to growth stage. But they can also inform interventions that are tailored to the local needs and situation.

In northern Ethiopia, the distribution and period of occurrence of rains (E_R and Ad respectively), the average daily temperature, GDD and KDD during vegetative growth stage influence yield variability. The presence of good predictability of above and below normal vegetative stage rainfall distribution and daily temperatures in lead months zero to two in North-Gonder provide good potential to predict maize yields.

Forecasts of rainfall indicators in North-Gonder during anthesis do not influence yields nor are they predictable. Anthesis period minimum temperature and daily temperature are predictable in but only lead-0 GDD. This is still useful because lead-0 refers to one month before sowing permitting possible prediction of yields and adequate time to make decisions before the crop reaches anthesis.

There is a significant correlation and predictability of reproductive stage daily temperature with three months lead-time in North-Gonder. Growing season rainfall characteristics forecasts in North-Gonder may be forecasted with certainty in all the three lead-times as opposed to temperature characteristics that show lower certainty in prediction.

Yield sensitivity to growing season temperature only seen in Bungoma may be due to both uniform and sufficient precipitation and soil moisture variability during the crop growing season i.e. there is less spatial and temporal variability as opposed to northern Ethiopia. In Bungoma however, yields show sensitivity to vegetative stage temperature indices only.

Anthesis rainfall indices in Bungoma do not explain yields probably due to the short period around anthesis. Anthesis period average minimum temperature and daily temperature are predictable in all the three lead-times in both North-Gonder and Bungoma. There is neither significant correlations nor sensitivity between reproductive stage rainfall, nor temperature indicators and yield variability in Bungoma.

Uncertainty in forecasts of climate indicators depend on region, maize growth phase but varies less with forecast lead-time. Importance of temperature and rainfall for maize growth varies in space, time and phenological stage.

While average climate conditions during a crops' growth cycle drives yields, this study has illustrated important localized, growth phase dependent yield sensitivities. This knowledge would enable yield forecasting without use of complex crop models and may provide a service by enabling tailored cheap, short duration, localized adaptation interventions to ensure good yields as opposed to generalized regional interventions.

Chapter 5

Probabilistic Maize Yield Prediction over East Africa using Dynamic Ensemble Seasonal Climate Forecasts

Abstract

We tested the usefulness of seasonal climate predictions for impacts prediction in eastern Africa. In regions where these seasonal predictions showed skill we tested if the skill also translated into maize yield forecasting skills. Using European Centre for Medium-Range Weather Forecasts (ECMWF) system-4 ensemble seasonal climate hindcasts for the period 1980-2011 different initialization dates before sowing, we generated a 15–member ensemble of yield predictions using the World Food Studies (WOFOST) crop model implemented for water-limited maize production and single season simulation. Maize yield predictions are validated against reference yield simulations using the WATCH Forcing Data ERA-Interim (WFDEI), focusing on the dominant sowing dates in the northern region (July), equatorial region (March-April) and in the southern region (December). These reference yields show good anomaly correlations compared to the official FAO and national reported statistics, but the average reference yield values are lower than those reported in Kenya and Ethiopia, but slightly higher in Tanzania. We use the ensemble mean, interannual variability, mean errors, Ranked Probability Skill Score (RPSS) and Relative Operating Curve skill Score (ROCSS) to assess regions of useful probabilistic prediction. Annual yield anomalies are predictable 2–months before sowing in most of the regions. Difference

This chapter has been published as:
Ogutu G.E., W.H. Franssen,I. Supit,P. Omondi and R.W. Hutjes,2018: Probabilistic maize yield prediction over East Africa using dynamic ensemble seasonal climate forecasts. *Agricultural and Forest Meteorology*, **250** , 243-261.

in interannual variability between the reference and predicted yields range from from ±40%, but higher interannual variability in predicted yield dominates. Anomaly correlations between the reference and predicted yields are largely positive and range from +0.3 to +0.6. The ROCSS illustrate good pre-season probabilistic prediction of above-normal and below–normal yields with at least 2–months lead time. From the sample sowing dates considered, we concluded that, there is potential to use dynamical seasonal climate forecasts with a process based crop simulation model WOFOST to predict anomalous water-limited maize yields.

5.1 Introduction

Agriculture is the major land use across the globe and is of high economic, social, and cultural importance. In its many forms, agriculture remains highly sensitive to both climate extremes and to variations in climate and trends on a range of time scales; particularly in regions where rainfed agriculture supports majority of the population and plays crucial roles in national economies like East Africa. Improving resilience of the agricultural sector by preparing the vulnerable populations for extreme weather variability and developing reliable crop production systems (Matthew et al., 2015) can not only have a positive effect on socio-economic development but also enhance food security through better agricultural management and policy formulation that proactively accounts for variable climatic conditions (Bahaga et al., 2016).

Operationally, efforts towards improved resilience to extreme climate variability are on-going through issuance of pre-season climate forecasts generated by both statistical and dynamical methods. In Eastern Africa, these forecasts are issued through the Greater Horn of Africa Climate Outlook Forum (GHACOFs) (Martinez et al., 2010; Ogallo et al., 2008) organized by the Intergovernmental Authority for Development (IGAD)- Climate Prediction and Applications Centre (ICPAC) and the World Meteorological Organization (WMO) together with other partners. It brings together scientists from the global climate producing centers, meteorologists from the National Meteorological and Hydrological Services (NMHS) from the GHA region, climate forecast end-users and the relevant stakeholders to develop a consensus rainfall and temperature forecasts for the coming season plus likely impacts on climate sensitive sectors (Hansen et al., 2011; Ogallo et al., 2008) including agriculture. The scientists further downscale the consensus seasonal climate outlooks for national impacts and other purposes. Seasonal climate impacts outlook are generally based on subjective expert judgement rather than explicit quantitative methods. Model based, quantitative pre-season crop yield forecasting plus communication of associated uncertainty and skill could be incorporated into GHACOF process to enhance use of seasonal climate forecasts by providing direct impacts on maize production, based on the assumption that predictable climate can be translated into predictability of crop phenological development and subsequent yields. This study presents the possibility of providing bio-physical process based, quantitative yield forecasts besides the seasonal climate forecasts already routinely issued.

A number of early warning systems (EWS) exist in East Africa with mandates to provide food security outlooks and warnings. For example, the United States Agency for International Development (USAID) Famine EWS (FEWS-NET) provides food security outlooks, assistance outlooks, markets and agricultural trading outlooks (Brown et al., 2007; Ververs, 2012). The Global Information and Early Warning Systems (GIEWS) of United Nations' Food and Agriculture Organization (FAO) (FAO, FAO; Ververs, 2012) provides information on crop prospects and food situation depending on emerging crisis often after crop and food security assessment missions. The Food Security and Nutrition Working Group (FSNWG), a regional platform whose members include NGOs, UN agencies, and research institutions, amongst others, provide food security and nutrition outlook in their monthly meetings. For crop monitoring, these organizations use agrometeorological assessment reports and satellite technologies that monitor conditions of food crops after planting for example the normalized difference vegetation index (NDVI), rainfall estimates, and expert judgement to estimate impending food security situations. The existing EWS largely focus on water availability without considering the water-temperature interactions, even though temperature is critical in both rainfed and irrigated agriculture as it influences the rate of crop development and water deficit in irrigated fields. The complex reactions between climate variables and crop physiology are better simulated using biophysical models as in in this study.

Since existing EWS monitor crops when they are already in the fields, little adaptation measures can be implemented to adjust to the prevailing climate situation. This study can directly expand the time horizon of crop performance prediction from existing EWS by including pre-season forecasts, and provide high resolution yield forecast information that is also relevant to farmers, rather than only to their traditional clients (i.e. governments and humanitarian agencies).

Seasonal climate forecasts are currently routinely issued up to 12-months before the start of seasons (lead-time) by numerous operational global forecast centers. With sufficient lead time before the start of a growing season, different adaptation options are possible (e.g. choosing different crops or varieties, heavy or low investment in farm inputs) as opposed to forecasts issued after crops are planted. Global Climate Model (GCM) based seasonal climate forecasts have been used in agricultural impacts modelling globally with varied results, suggesting variations in skill due to factors like spatio-temporal scales used, level of surface heterogeneity, crop management practices, and model initialization, amongst others (Jones et al., 2000; Lawless and Semenov, 2005; Neumann et al., 2010; Shin et al., 2010). Driving crop models with skillful seasonal climate forecasts may not guarantee good yield forecasts (Baigorria et al., 2007; Semenov and Doblas-Reyes, 2007; Shin et al., 2010), but the reverse, i.e. better skill in the crop forecast than in the meteorological forecast has also been reported (McIntosh et al., 2005). In addition, whether a crop in a certain region experiences temperature or moisture limitations affects yield predictability differently. For example, since temperature influences crop phonological development and its predictability is generally higher than for precipitation (Iizumi et al., 2013; Ogutu et al., 2017), its predictability influences yield predictability differently. Finally, the time of the year in which a forecast is useful depends on the crop and region (McIntosh et al., 2007), i.e. depends on the local cropping calendars. This study seeks to identify lead times and regions in East Africa with useful preseason yield predictability based on pre-season climate forecasts.

Seasonal crop yield forecasts have been derived from either historical statistical relationships with rainfall or large scale climate indices such as the El Nino Southern Oscillation (ENSO) Index (Amissah-Arthur et al., 2002; Iizumi et al., 2014; Hansen et al., 2009; Martin et al., 2000; Phillips et al., 1998), and its influence on seasonal rainfall in some parts of the world such as eastern and southern Africa. These statistical methods are successful at broader spatial extents like national boundaries or regions (Amissah-Arthur et al., 2002; Iizumi et al., 2014; Lobell and Burke, 2010; Phillips et al., 1998; Thornton et al., 2009) and may not suffice for smaller spatial scales where heterogeneities exist. For example, above normal rainfall season may result in low yields related to nutrient leaching depending on soil types. High rainfall variability exist in small regional extents even in an otherwise "good rainfall seasons" and statistical relationships do not capture rainfall characteristics (such as distribution during a season and frequency) that are important for crop yields. Weaknesses related to the use of large scale climate indices to forecast yields are highlighted in Mjelde and Keplinger (1998). Poor records of historical yields on which the statistical models are calibrated also influence prediction skill.

Confronted with the current climate change and variability together with climate teleconnections between a region of interest and other parts of the globe, any past statistical relationships between yields and climate indices may no longer hold true because the future will be under climate regimes (variability) not observed before. It is not clear if the relationships between phenological observations and satellite derived vegetation indices will hold true since observations will also be under different climate regimes (for example higher temperatures than in the historical period) and since crop response to climate is not linear (Porter and Semenov, 2005), mean historical observations may not suffice. Most studies related to yield impacts modelling over East Africa use GCM outputs to assess future climate change impacts on yields. In this study, we explore the use of seasonal forecasts and crop models to simulate yields at the shorter seasonal scales that determine year-to-year food production.

This work explores the use of dynamical seasonal climate forecasts based on Global Climate Models to simulate agricultural impacts. We assess ensemble (probabilistic) predictive skill of maize yields based on GCM seasonal climate forecasts via both baseline and hindcasts validation for the period 1981-2010. The aim is to identify lead times and areas of potential pre-season yield forecasting based on seasonal climate forecasts and maize planting dates. We assess how well yield forecasts capture observed/reference yield anomalies due to interannual climate variability and climate anomalies.

Because of inherent biases in climate models, bias correction of model output is important for impact studies. For example, biases in temperature would grossly affect simulation of maize phenology which depends on (cumulative) thermal time units during growing period. This study therefore uses bias corrected climate forecasts.

5.2 Materials and methods

5.2.1 Model description

Hindcast grid point maize yield forecasts over East Africa are simulated using the World Food Studies crop simulation model (WOFOST); a simulation model for the quantitative analysis of the growth and production of annual crops. WOFOST is a detailed model with respect to crop physiology allowing for example a specification of regionally used varieties. It was originally developed to simulate crop yield for a single location with homogeneous cultivars, soil characteristics and weather (Boogaard et al., 2013; Van Diepen et al., 1989; Keulen and Wolf, 1986; Supit et al., 2010) but is adopted for use in a larger region in this study. It has been used widely for many studies such as climate change impacts (Supit et al., 2012, 2010), regional yield forecasting and crop yield analysis in Ethiopia and Kenya (Hengsdijk et al., 2005; Rotter, 2014).

The model is photosynthetically driven and simulates the growth and production of annual crops using a range of physiological processes at daily time steps from sowing to maturity in response to weather, soil type, soil moisture conditions and as defined by crop characteristics. The physical processes simulated include:- light interception, photosynthesis and respiration, evapotranspiration, assimilate partitioning, leaf area dynamics, phenological development and root growth determined by cultivar characteristics, soil water balance and drought response.

We use the World Food Studies crop simulation model (WOFOST) version in the Python Crop Simulation Environment (PCSE/WOFOST) implementation (de Wit et al., 2019) to simulate maize yield under rainfed conditions. Details on this implentation are available at https://pcse.readthedocs.io/en/stable/index.html. In this study, the model simulates theoretical water-limited-yields (WLY) at simulation units of approximately $0.1^{\circ} \times 01^{\circ}$ resolution determined by overlaying land use, elevation, crop calendar, soil maps and climate data. The yields are then aggregated to 0.5° grid cells (corresponding to climate data grid) weighted by cultivated area within each grid (Boogaard et al., 2013). WLY is influenced by rainfall assuming a soil nitrogen level that reflects a maximum yield level of about 40-50% of the potential, typical for the region, and otherwise optimal crop management and that no losses occur due to pests or diseases. Only the main cropping season is modelled from emergence to maturity; this may result in lower than observed annual national yields (as reported in e.g. FAO and national reported agricultural statistics), because double-season cropping is not implemented in areas where it exists (e.g. Belg rains receiving areas of Ethiopia, eastern and northern Ethiopia, and the bimodal rains region of Kenya and Tanzania (see Supplementary Figure C.10). The model has been calibrated, applied and validated for maize production in Kenya (Rotter, 2014); in the Central Rift Valley of Ethiopia (Kassie et al., 2014, 2015); and in Tanzania in the Global Yield Gap Atlas project (Makio, 2020). Further details on WOFOST are also available at http://www. supit.net/ and http://www.wageningenur.nl/en/Expertise-Services/Research-Institutes/alterra/Facilities-Products/Software-and-models/WOF OST/Implementation-WOFOST.htm.

5.2.2 WOFOST input data

5.2.2.1 Sowing dates and crop varieties

Crop sowing dates impact accurate simulation of crop growth and yields. A range of approaches that include: use of observed sowing dates; use of rainfall characteristics at the start of a season; optimization of dates on crop-and-climate specific characteristics; optimization of sowing dates on maximum yields, amongst others, have been used to setup sowing dates in simulation studies (Srivastava et al., 2016). In this study, we have used a method that combines the "optimization of sowing dates on maximum yields" and "optimization on climate and crop specific characteristics". This procedure has been used and detailed in (Wolf et al., 2015).

Briefly, we start with coarse crop calendars from FAO (http://www.fao.org/agriculture /seed/cropcalendar/welcome.do) and Sacks' calendar (Sacks et al., 2010), and standard tropical maize varieties for crop modelling compiled in Van Diepen et al. (1988) and Van Diepen et al. (1988). Yields are simulated for the different maize varieties planted at every grid point every 10 days, starting 90 days before and end 90 days after the sowing dates presented in the coarse crop calendars, using the baseline climate forcing for the period 1981 to 2010. Maize varieties differ mainly in thermal time accumulation needed to reach flowering and maturity. Thermal time accumulation in WOFOST is based on the planting dates and a base temperature of 10°C. We subsequently selected the sowing date and crop variety that provided highest crop yield averaged over the baseline period, resulting in 10 varieties in East Africa (see Table-5.1). The derived sowing date for each grid point was fixed in WOFOST, while harvest dates are determined by the weather conditions during growth period. The same sowing dates are used for both the baseline and forecast yield simulations. Note that in reality there are more than 2000 maize varieties in East Africa and sowing date may also be influenced by other, non-climatic factors such as availability of labour, seeds, customs and traditions etc.

5.2.2.2 Land use and soil data

Land use and soil input data are derived from the International Soil Reference and Information Centre-World Soil Information (ISRIC-WISE) database Batjes (2012) and http://www.isric.org/data/isric-wise-international-soil-profile-dataset). The database includes information on soil physical characteristics, root depth, and landscape characteristics such as elevation, slope gradients, and slope aspects. Soil properties such as wilting point and field capacity are estimated with the pedotransfer functions from Saxton et al. (1986). Maize growing areas are determined based on FAO land use maps (Fischer et al., 2008). To ensure germination, the threshold soil water content was set to 10mm. Derived nominal crop varieties based on optimized TSUMs

Table 5.1: Derived nominal crop varieties based on optimized TSUMs. *Note:TSUM1 = temperature sum from emergence to anthesis [degree days]; TSUM2 = temperature sum from anthesis to maturity [degree days]*

MAIZE VARIETY	PARAMETER NAME	VALUE
1	TSUM1	400
	TSUM2	350
2	TSUM1	550
	TSUM2	500
3	TSUM1	620
	TSUM2	570
4	TSUM1	670
	TSUM2	620
5	TSUM1	720
	TSUM2	670
6	TSUM1	770
	TSUM2	720
7	TSUM1	820
	TSUM2	770
8	TSUM1	870
	TSUM2	820
9	TSUM1	920
	TSUM2	870
10	TSUM1	970
	TSUM2	920

5.2.2.3 Weather data

The sparse climatological station network in East Africa limits the use of pure observed meteorological station data to simulate crop yields over a large area. Moreover, station data are impractical for the evaluation of gridded products, even though they are invaluable for farm level or field experimental studies. We therefore use requisite weather data (i.e. precipitation (*tp*), maximum and minimum temperatures (*tmax* and *tmin* respectively), surface downward shortwave radiation (*rsds*) from a model, and an observation fused data product from the Water and Global Change (WATCH) forcing data ERA-Interim at $0.5^{\circ} \times 0.5^{\circ}$ resolution (Harding et al., 2011; Weedon et al., 2014, 2011, 2010) to simulate grid point reference yields. WFDEI is obtained after elevation correction and monthly bias correction of the ERA-Interim re-analysis (Dee et al., 2011) using gridded observed climate data from University of East Anglia's Climate Research Unit (CRU)(Harris et al., 2014). ERA-Interim is one of several available climate reanalysis products providing a numerical description of the recent climate; produced by assimilation of weather observations in forecast model simulations. Reanalysis are considered the most consistent multivariate representation of the past climate.

In this study we use WFDEI to represent the observed climate during the study period.

To simulate yield forecasts, we use climate reforecasts (also called hindcasts) from the ECMWF System-4 seasonal climate ensemble pre-diction system (Molteni et al., 2011)), and bias corrected against WFDEI. Seasonal climate forecasts are initialized on the first day of every month from 1981 to 2010 with 15–perturbed (different) initial conditions. Since weather evolution during an upcoming season is not certain, initializing the climate forecast model from different initial conditions allows sampling of the possible weather evolution path during the coming season. Forecasts initialized from each of the perturbed initial conditions results in 15-different forecasts (ensemble members). Each member provides a forecast 7–months into the future. Details of initial conditions generation (perturbations) and forecast ensemble generation can be obtained from Molteni et al. (2011).

Useful probabilistic prediction skill in climate forecast exists in all major cropping regions of East Africa, but the skill level decreases from near surface air temperature; to precipitation; to surface downwelling surface radiation. A detailed analysis of applied climate forecast for the region is presented in (Ogutu et al., 2017). To simulate yield forecasts, WOFOST is driven with each of the 15 climate forecast ensemble members. This results in 15–ensemble yield forecasts, i.e. it spans the range of possible yields resulting from the 15 probable evolution of climate during the crop growing period.

5.2.3 Validation methodology

Our validation and verification methodology is outlined in Figure-5.1. We generate baseline or reference yields from WOFOST driven by WFDEI climate (henceforth referred to as WFDEI yields). WFDEI yields are compared to country level yield statistics from the Food and Agriculture Organization yield statistics (FAO) and to nationally reported yields (NAT) to evaluate how well the WFDEI yields corresponds to the "observed". The national annual yield statistics are estimates of actual yields from agriculture ministry or authorities in each country obtained by agricultural census or farmer interviews in sub-national districts and aggregated to country levels. The surveys are carried out by the individual countries, as such; survey methodologies may not be uniform. Observed yields are not expected to represent scientific biophysical relationships between climate variables and crop characteristics as represented in WFDEI yield simulations. Further, yield data from agricultural surveys may depend on socio-economic factors not incorporated in WOFOST. Our interest is to simulate climate related impacts on yields and will therefore focus on yield anomalies rather than actual values. Anomalies for each data set are calculated by taking the difference between annual yield and detrended mean yields during the study period, and normalised by the standard deviation over the same period.

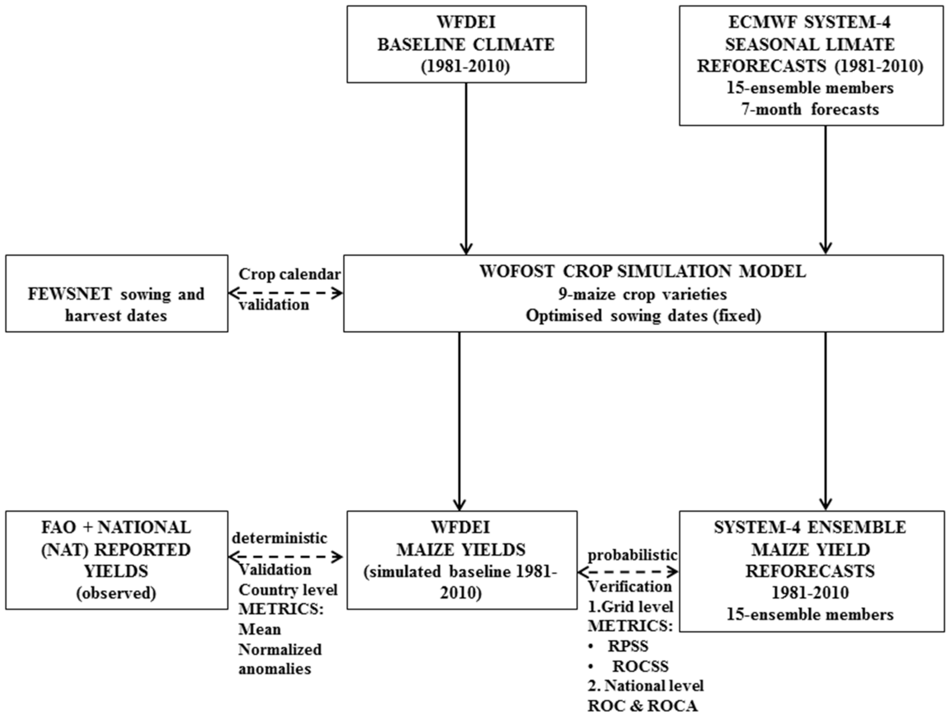


Figure 5.1: Study setup and validation methodology

We use WFDEI yields to validate yield reforecasts because gridded observed yields data are non-existent for East Africa (or elsewhere). Harvested yields from all grid points planted in a chosen month are used to assess yield predictability for particular lead-times (months before sowing). Because each climate forecast runs for seven months ahead from the initialization month we use forecasts that allow growth from sowing to maturity (i.e. a complete growing cycle). The weight in storage organs achieved during the growing period from planting to harvest is taken as the yield forecast.

'Lead months' correspond to the number of months before sowing that a climate forecast was initialized. For example, if maize is sown in March, forecasts initialized in March will represent lead-0, forecasts initialized in February represents lead-1 and so forth, and this applies to both climate and yield forecasts.

Validation is performed on yields aggregated to national country boundaries, and grid-point level corresponding to climate grid. We use coefficient of variation (CV) measure to assess the interannual variability of the reference and forecast yields. To verify ensemble predictions we use probabilistic metrics that are based on contingency tables of hits and false alarms of predicted and observed yield falling in a certain percentile.

We use the ranked probability skill score (RPSS) to examine the average model skill and the relative operating characteristic curve skill score (ROCSS) to assess skill of yield prediction in specific terciles, i.e. the near-normal (middle tercile), below-normal (lower) and the above-normal (upper) yields, respectively. Below-normal (BN) and above-normal (AN) yields are below the 0.33 and above the 0.66 percentiles of the historical yields respectively. Near-normal (NN) yield fall in between these 0.33 and 0.66 percentiles. A set of hit-rates and false alarm rates constructed from the ensemble of reforecasted yields are determined and plotted for a range of decile forecast probability thresholds for each forecast category (AN, NN, or BN) thus generating a ROC curve (e.g. see Figure 5.7). Hit-rates refer to the proportion of events that occur as forecasted; while the false-alarm rates are the proportion of events forecasted but do not occur. The area under the ROC curve (AROC) indicates the performance of the forecast. Because there is skill only when the hit rates are higher than the false-alarm rates, the ROC curve lies above the 45-degree line (1:1 line) for a skilful forecast (i.e. AROC > 0.5), and below (AROC < 0.5) for a forecast with no skill (see Figure 5.7 and supplementary Figure C.12). A score of 1.0 (AROC $= 1$) indicates a perfect forecast while AROC $= 0.5$ indicates the forecast is as good as a climatological forecast. P-value determined from Mann-Whitney U statistics show the significance of AROC i.e. p-values < 0.05 show significant scores at 95% significance level. This measure is detailed in Barnston et al. (2010); Buizza and Palmer (1998); Diro et al. (2012); Hamill and Juras (2006); Mason and Graham (1999)See Appendix B.2 for a detailed description of RPSS.

5.3 Results

We start by comparing WOFOST determined sowing dates and simulated harvest dates to known cropping calendars for the region from the Famine Early Warning System Network (FEWSNET;http://www. fews.net/east-africa/kenya/seasonal-calendar, http://www.fews.net/ east-africa /ethiopia/seasonal-calendar, and http:// www.fews.net/ east-africa/tanzania/seasonal-calendar) shown in supplementary Figure–C.10. We compare annual WFDEI yields to both FAO and NAT statistics to assess model simulation accuracy. For a useful yield forecasting system we do not necessarily need equality in mean yields, as already mentioned, rather we need a level of similarity of yield anomalies as influenced by climate anomalies. We then describe results of deterministic and probabilistic validation of yield forecasts at both the national (country) and $0.5^{\circ}\times0.5^{\circ}$ grid point levels.

5.3.1 Crop calendars

WFDEI determined reference sowing and harvest dates are shown in Figures 5.2(a) and (b) respectively. Reforecasted harvest dates averaged over fifteen ensemble members and forecast lead times are in Figure 5.2(c). We compare these simulated calendars to the FEWSNET calendars (see supplementary Figure C.10).

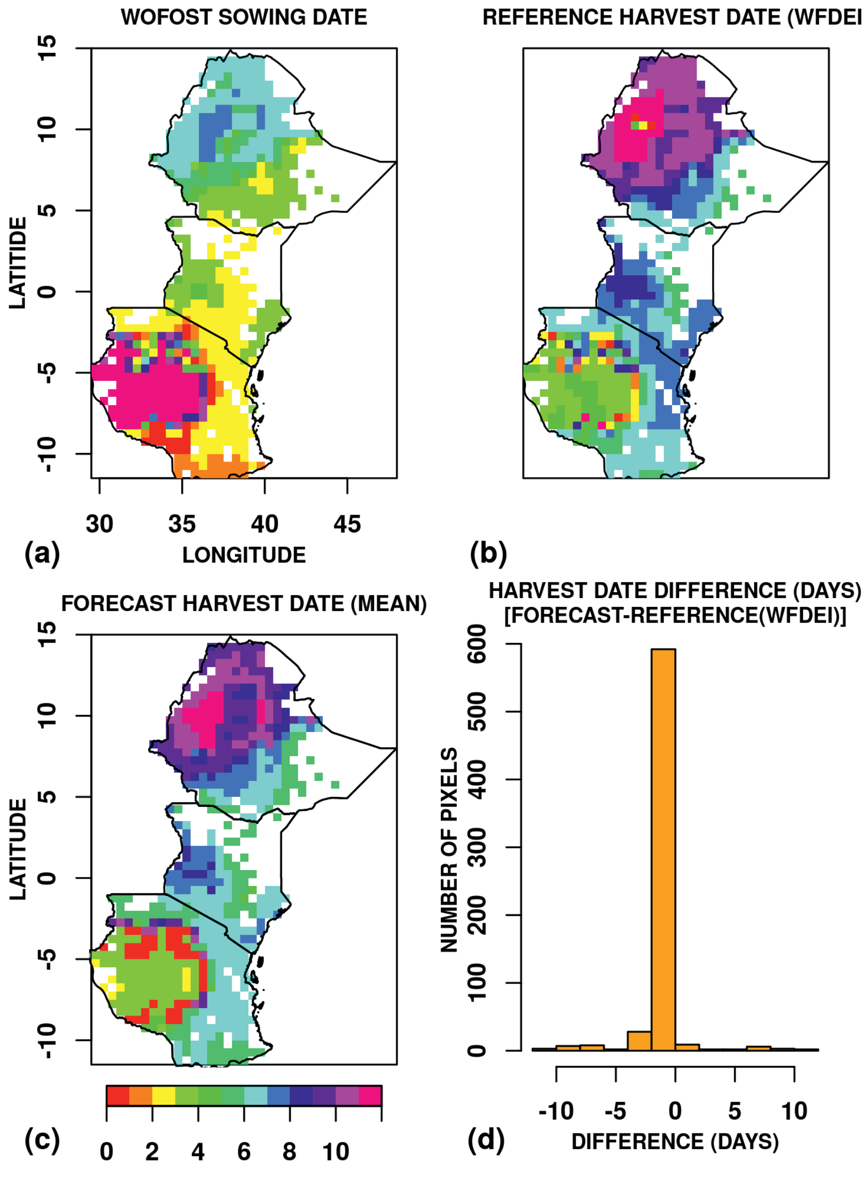


Figure 5.2: Figure of study area showing the reference mean sowing dates (WFDEI) (a), reference harvest dates (b), average reforecasted harvest dates (c) and the difference between WFDEI harvest dates and reforecasted harvest dates across 15-ensemble members (d). Dates shown in months of the year.

WFDEI derived sowing dates (5.2(a)) and simulated reference and forecasted harvest dates (5.2(b) and (c)) correspond to FEWSNET calendars accurately. But because it is configured for only one season in a year, some planting seasons for example the short rains planting season in Kenya is not simulated. Harvest dates in both the reference and hindcast yield forecasts in our study are fairly replicated with differences being less than 10 days (5.2(d)). In a majority of pixels, the differences are near zero. We conclude that the method used to determine sowing dates is accurate and suitable for our study, and the simulated harvest dates correspond well to the typical crop calendars for the region.

5.3.2 Deterministic validation

5.3.2.1 National annual average simulated grain yields, anomalies, and interannual variability

We compare our reference WFDEI yields to both NAT and FAO for the period 1981–2010. Figure 5.3 and 5.4 respectively show box plots and interannual anomaly time series of WFDEI yields, FAO (1981–2010 for Kenya and Tanzania; 1993–2010 for Ethiopia) and NAT (1981–2005 for Kenya; 1997–2010 for Tanzania; 1995-2008 for Ethiopia) yields.

The difference in mean (and median) yields between WFDEI, FAO and NAT are underestimated by between 600–700Kg ha^{-1} by the model over Kenya and Ethiopia, but less than 200Kg ha^{-1} over-estimation by the model over Tanzania. Tanzania FAO yield anomalies show a remarkable difference from the simulated reference yields, (Figure 5.4b) characterized with high yields from 1995 to 2002. These outliers are unique only to FAO statistics and cannot be explained either from the NAT or WFDEI yields. Further, differences in FAO and NAT statistics also highlight the uncertainty in official observed yields which are not surprising as already highlighted in Section 5.2. The differences described are also visualized in Figure 5.3a to 5.3c for Kenya Tanzania and Ethiopia respectively. FAO yield distributions are quite different from NAT and WFDEI in Tanzania and Ethiopia. Figure 5.4 compares anomaly time series of WFDEI, FAO and NAT yields. Some yield anomalies are fairly well captured in the three datasets, notably for Kenya (Figure 5.4a). Less variability in Tanzania WFDEI yields (Figure 5.4b) results in less anomalous yields especially in the period 1997–2010 compared to FAO. Yield anomalies in Figures 5.4 and 5.5 subjectively correspond well to the interannual rainfall anomalies (shown in Figure 5.4) indicating the crucial role climate plays in maize production or simulation even though low correlations between WFDEI precipitation and both FAO and NAT are apparent (i.e. $r = 0.02$ to 0.2 against FAO and -0.15 to 0.2 against NAT). Reason being that rainfall (or climate in general) is not the only factor that influences actual yields. Anomaly correlation coefficient between annual WFDEI rainfall anomalies and simulated WFDEI reference yield anomalies show good correlations of 0.4 to 0.6.

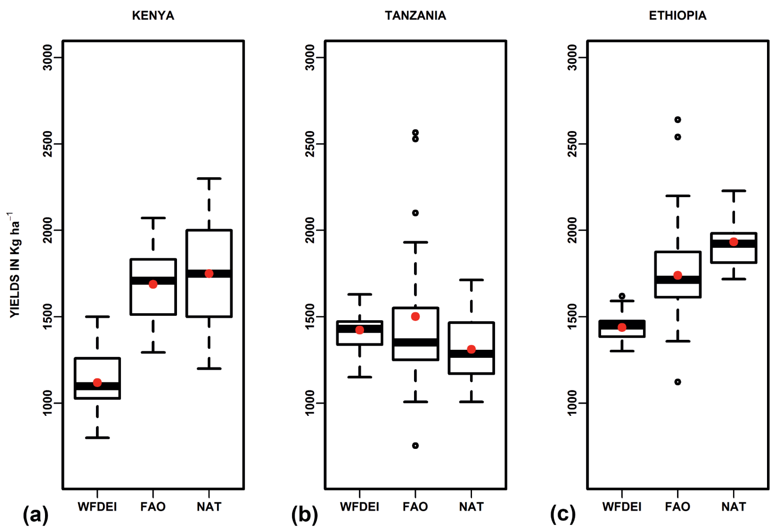


Figure 5.3: Box plots showing comparison of mean and data range of WFDEI yields, FAO yield and national (NAT) reported yields for Kenya, Tanzania, and Ethiopia. Thick line in the centre of the box represent the median yield; mean yield is represented by red dots; top of the box represent the 75^{th} percentile; the bottom of the box are the maximum and minimum values respectively while the length of the box is the interquartile range (IQR). Suspected outliers are yields 1.5×IQR or more above the third quartile or 1.5×IQR or more below the first quartile. Note the the three data sets do not span the same period as shown in Figure 5.4.

5.3.2.2 National yield predictions

We assess the prediction of annual mean yields by comparing WFDEI yields to forecasted yields for each country focusing on chosen sowing dates. Predictions in form of standardized anomalies are shown in Figure 5.5 as time series of ensemble average forecasted yields at lead-times 0 to 3 months before sowing. Predictability of yield anomalies varies with each sowing date and region. Some anomalies are commonly captured and potentially predicted at least 2-months before sowing over the entire region for example 1984, 1989, 1997, and 1998. These years correspond to extreme El-Niño Southern Oscillation (ENSO) events illustrating the influence of ENSO on crop production. Predictability of other annual yield anomalies depend on sowing dates for example, predictability of harvests of maize planted in April in Kenya and Ethiopia (Figure 5.5a and b) are better than harvests for maize planted in March (Figure 5.5e) and July (Figure 5.5d). Likewise predictability and even variance of Tanzania harvests sowed in December and March (Figure 5.5c and f respectively) differ. In general though, anomalous annual yields over East Africa are predictable 3-lead months before sowing except in Tanzania. This could be related to less variance of yields in Tanzania during the study period, which in turn may be related to climate characteristics. The inter-tropical convergence zone (ITCZ) which is responsible for spatial distribution of rainfall lies over Tanzania from October–April/May and may be responsible for less variability in climate during growing season and hence less variability in maize yields.

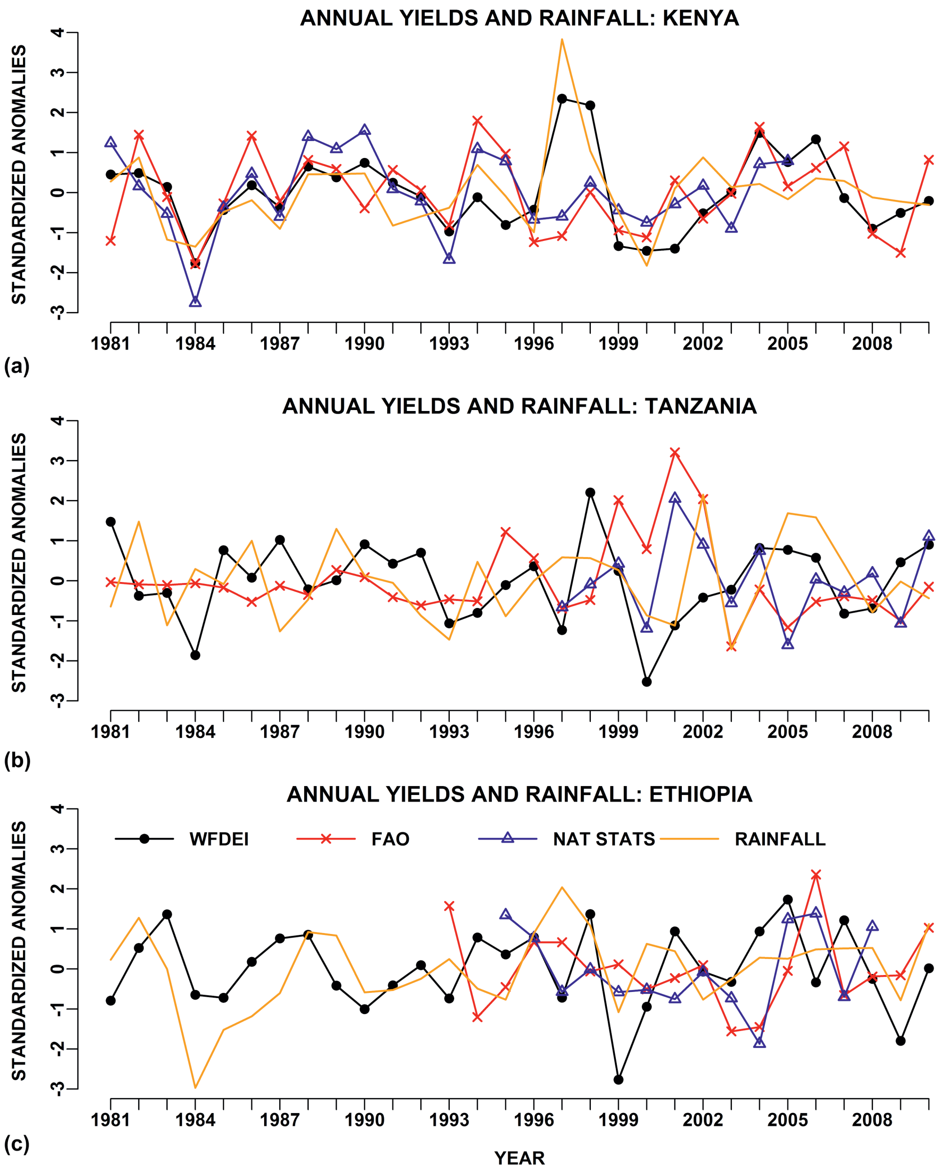


Figure 5.4: Time series of WFDEI, FAO, and NAT standardized yield anomalies plus standardised annual rainfall anomalies (a-c) for Kenya, Tanzania and Ethiopia.

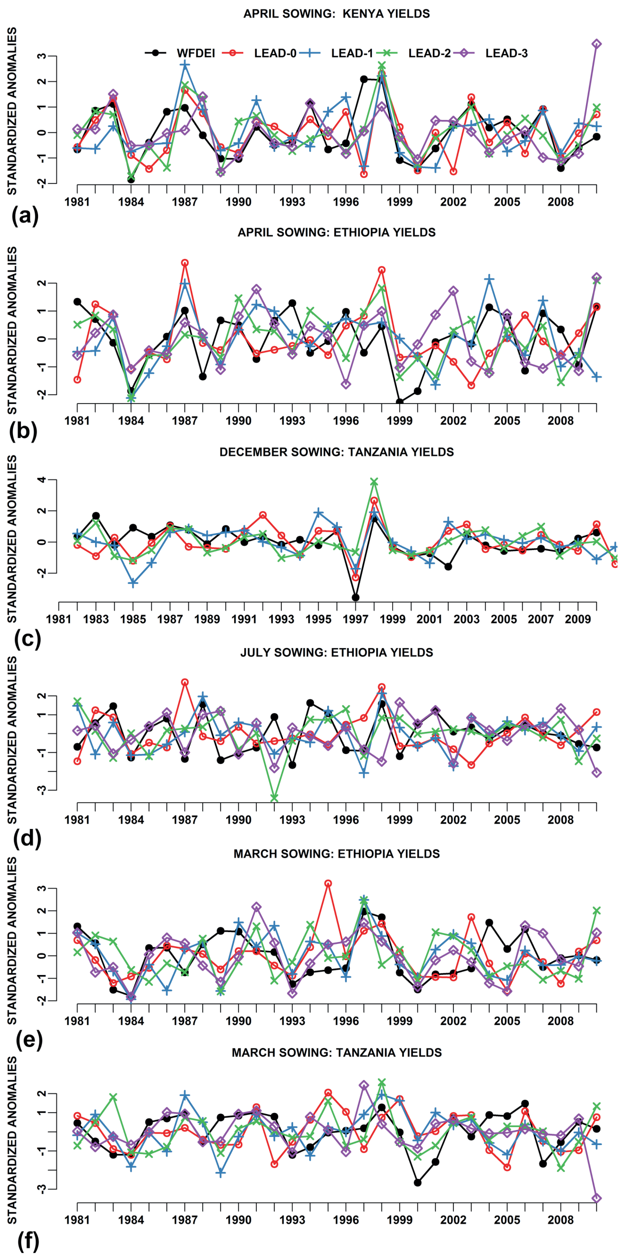


Figure 5.5: Predicted yield anomalies (standardized) and WFDEI yield anomalies (standardized) shown for particular dominant sowing dates in each country and forecast lead times. Time series plot show interannual variability of the reference and reforecasted yields for varying forecast lead months.

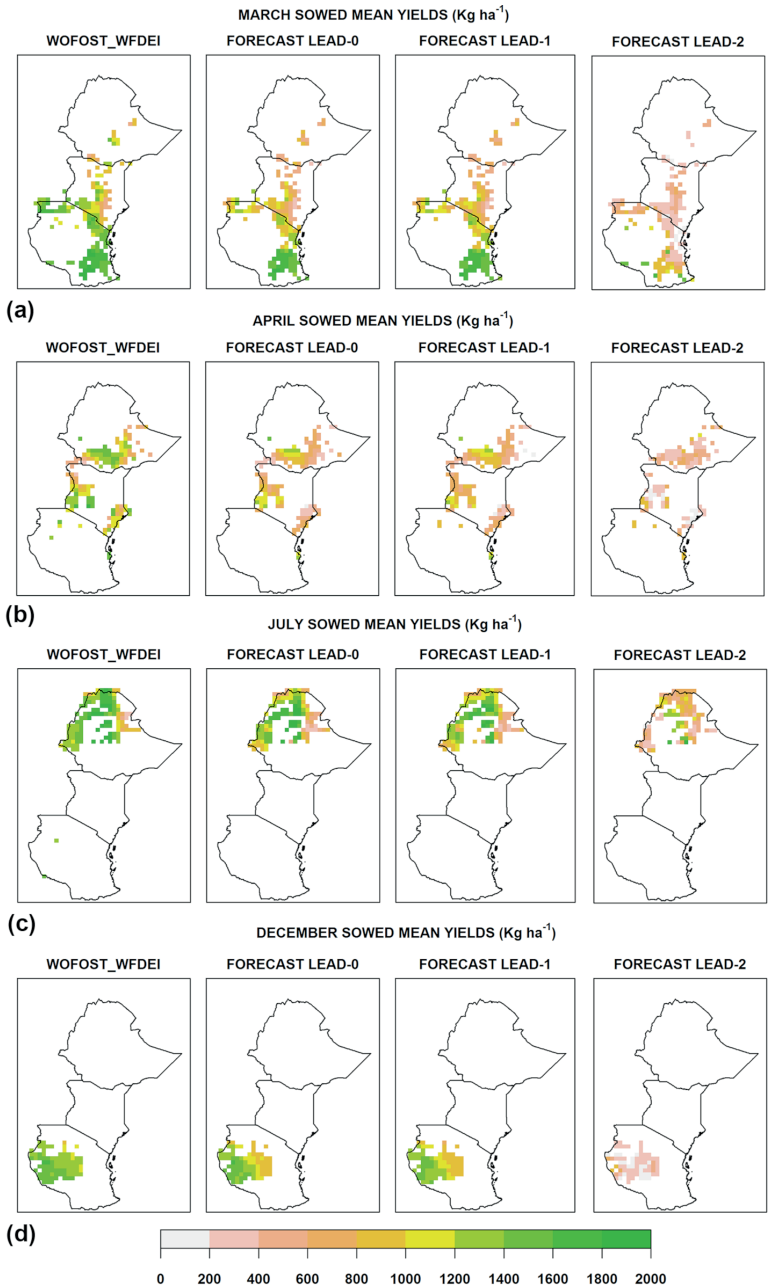


Figure 5.6: Mean yield for the period 1981-2010 (in $Kgha^{-1}$) for reference and reforecast yields at various sowing dates.

5.3.3 Grid point validation

In this section, we use a range of deterministic and probabilistic measures to compare WFDEI and forecasted yields for different sowing dates and forecast lead-times. Maps are used to show the spatial distribution of both deterministic and probabilistic (i.e. ROCSS and RPSS) verification measures.

5.3.3.1 Average yields, deviation and interannual variability

Spatial patterns of WFDEI and forecasted yields are similar (Figure 5.6). Differences in WFDEI and forecasted mean yields depend on sowing dates, forecast lead-time and region of the study area. Mean forecasted yields reduce with increasing lead-time before sowing dates. This may be because maize does not grow to maturity during the 7-month forecast period in some grids (i.e. fewer grids are harvested at long lead-months). Large deviations from mean yields indicate areas of large spread of simulations and may be a result of the interannual variability of climate. We therefore show this variability using the coefficient of variation (CV) in the supplementary material Figure C.11. The forecasted yield coefficient of variation is the average of the CVs from each ensemble member (i.e. average of 30 years $\times$ 15-ensemble members). Higher (lower) percentages indicate higher (lower) variability in yields. Subtracting forecasted yield CV from WFDEI yields CV, we show areas where inter-annual variability in the reference yields is higher or lower (Figure C.11). Positive CV (red colours) implies higher inter-annual variability in WFDEI yields. In many grid cells, forecasted yields generally show higher inter-annual variability than the WFDEI (i.e. CV maximum difference is in the range $\pm 40\%$ but differences less than $\pm 20\%$ dominates). The magnitudes however depend on the sowing dates. March and April sowing dates show a number of grid cells where WFDEI yields have higher inter-annual variability mainly in drier regions suggesting a dependency of yields on climate. In dry areas, yields are more dependent on the availability of rainfall hence large differences (and mean deviation) during years of rainfall and in years of drought. The least interannual variability is associated with December sowing dates in Tanzania, probable reasons for this is highlighted in Section 5.3.2. Even though there are differences in inter-annual variability, the mean errors between WFDEI and mean reforecasted yields are low, not exceeding $200\mathrm{Kg\,ha^{-1}}$. In general forecasted mean yields approximate the reference yields, the difference in interannual variability are no more than 40% implying potentially good yield predictions.

5.3.3.2 Probabilistic validation

We present the predictability of above-normal (AN), below-normal (BN) and near-normal (NN) terciles relative to climatological yield forecasts using area under Relative Operating Curve (AROC) and it skill sore (ROCSS). We use the Ranked Probability Skill Score (RPSS) to show the average skill in forecasting all yield forecast probabilities relative to WFDEI yield climatology at each $0.5^{\circ} \times 0.5^{\circ}$ grid cell.

Climatological yield forecasts are taken as the average of reforecasted yield over all the 15-members.

We show the skill of prediction of national yields as a function of lead-time using ROC curve, AROC and its significance (p-value) in Figure 5.7 and supplementary Figure C.12 for sample sowing dates. Lead time of significant skill depends on sowing date and forecast tercile; for example, Ethiopian AN yields planted in July show significant skill with 1-month lead-time (AROC = 0.68 to 0.86) while BN show significant skill with 2-months lead time (Figure 5.7b). Kenyan yields planted in April (Figure 5.7a) show above normal and below normal significant skill at lead-months 0, 2 and 3 but insignificant at lead-1. This may be related to less skill in lead-1 climate forecasts also found and discussed in Ogutu et al. (2017). Other sowing dates shown in supplementary Figure C.12 show less skill.

We convert AROC into the relative operating curve skill score (ROCSS). The percentage of planted grids with significant skill for each sowing date is shown in Figure 5.8. There are apparently few grids planted in December with significant ROCSS (Figure 5.8d), this is even less than the 5% of grid cells to be expected to show skill by chance (at the 95% significance level used here). This illustrates the lack of skill in forecasting December planted yield terciles. Skill at 2-months lead-time could be related to rainfall forecast skill also seen at longer lead times in the October-December and March-May rainfall seasons (Ogutu et al., 2017). The proportion of planted grids with significant skill scores generally varies with forecast lead-time, sowing dates and regions. We cannot say whether AN yield forecast skill is systematically higher than for BN, nor vice versa but both are higher than NN yield forecast skill. This feature is also seen in probabilistic climate forecasts.

We use spatial maps to show distribution of ROCSS for March, April, July and December sowing dates in Figures 5.9 to 5.12 respectively. In all the figures, dotted cells indicate areas where the scores are significantly greater than zero at 95% confidence level. The percentage of cells with significant forecast skill relative to planted cells is already shown in Figure 5.8 and discussed in the above paragraph. In Figures 5.9, 5.10, 5.11 and 5.12, ROCSS show that above normal (AN) and below normal (BN) yields are predictable compared to near-normal forecasts. In Figures 5.9 and 5.10, i.e. yields planted in March and April respectively, many cells have good predictability of BN and AN yield forecasts three months before sowing date. Similar characteristic is shown in Figure 5.11, i.e. July sowing date in Ethiopia. This sowing date also shows higher number of cells with good predictability compared to March and April sowing dates (5.9 and 5.10 respectively) Maize planted in December i.e. over Tanzania show no potential predictability.

Grid point ranked probability skill score (RPSS) for the same sowing dates are in supplementary Figure C.14. RPSS being an average skill measure over all yield forecast categories (decile categories) exhibit lower skill than seen in ROCSS.

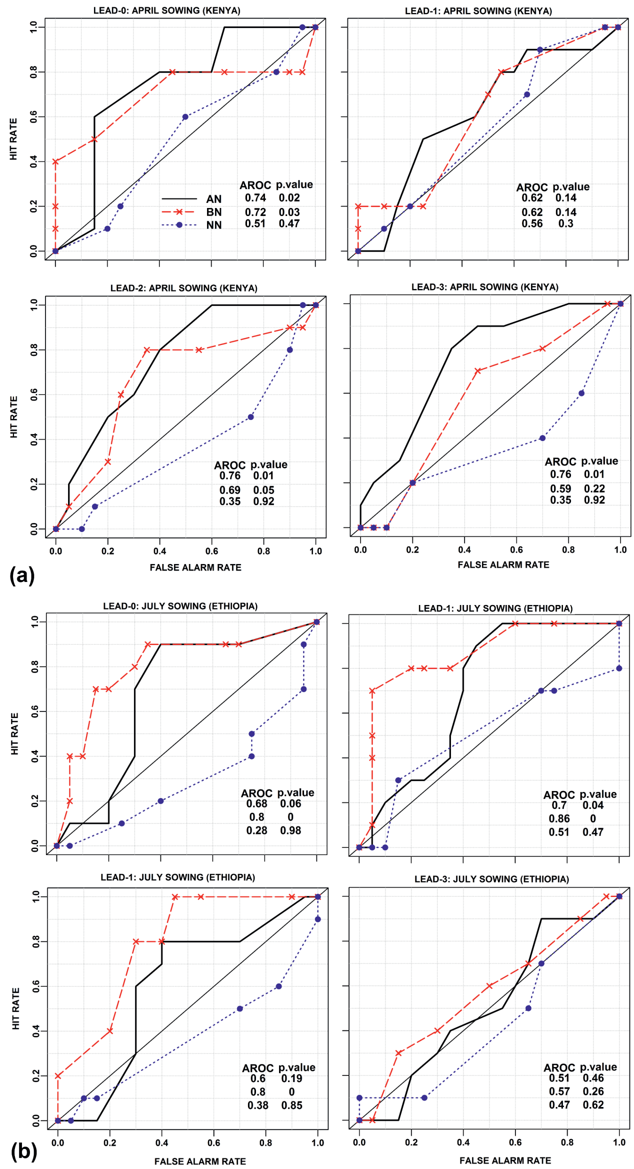


Figure 5.7: ROC curves of April (Kenya) and July (Ethiopia) reforecasted yields (a and b respectively). Validation is done at country level to show skill of tercile prediction of above-normal (AN), below-normal (BN) and near-normal (NN) yields. Area under ROC (AROC) and significance level for these 3-categories are given in the inset. Other sowing dates are in supplementary Figures C.12 and C.13

Very few grids show significant RPSS scores with the exception of July sowing date over Ethiopia (supplementary Figure C.14b) where more grid cells show significant scores. In general, the RPSS are quite low and may not demonstrate the potential prediction skill that actually exists. Good ROCSS skill illustrates the potential to offer good anomalous above or below normal yield forecasts from seasonal climate forecasts with at least 2-lead months.

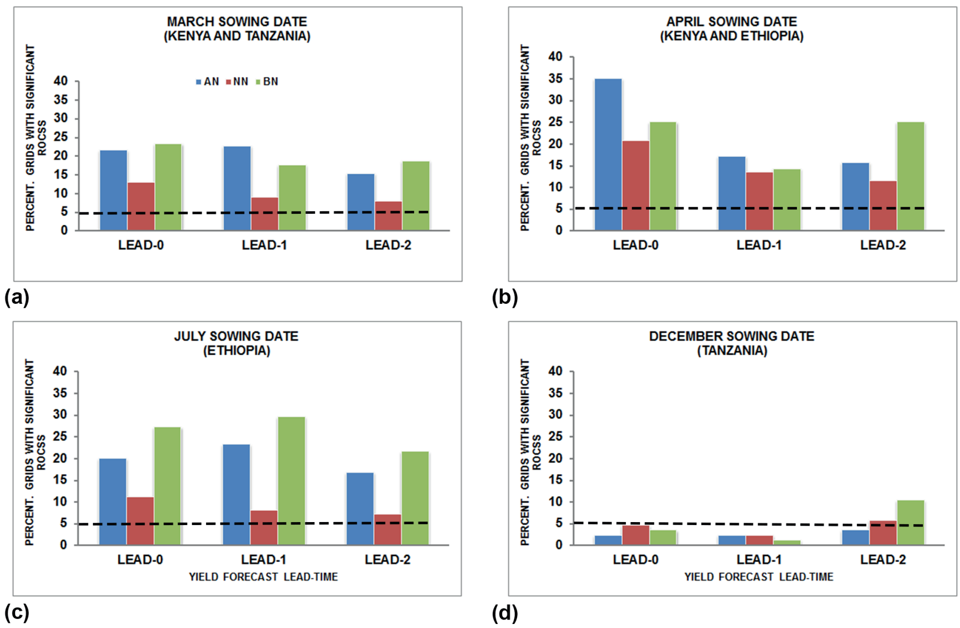


Figure 5.8: Percentage of grid points with significant ROCSS (at 95% level) fro the relevant sowing dates and forecast lead-times. The horizontal dashed line at 5% level shows a threshold below which skill may be purely by chance.

5.4 Discussion

5.4.1 Regional sowing dates

Grid based planting dates as established in this study allow crop models to be used in areas where high resolution crop calendars are not available or where the available calendars are too coarse such as in Sacks et al. (2010). We optimized sowing dates on a combination of maximum yields and climate-crop specific characteristics (Srivastava et al., 2016) resulting from an interaction of cultivars, soil properties, and climate in each crop simulation unit thus providing the best possible dates for this study. The method used samples various planting dates, this also samples the uncertainty space related to planting dates. Over regions of high climate variability and topographical variations, such as East Africa, setting dates for each grid point or simulation unit is more appropriate than using the available coarse generic calendars. It would also be futile to fit actual observed sowing dates to each grid cell because of high variability. A similar methodology to derive sowing dates i.e. running the model for a number of days around an assumed sowing date, have resulted in the best dates for highest water limited yields over Burkina Faso (Wolf et al., 2015). This lends credibility to apply a similar methodology in the present study. The method also agrees with findings in Stehfest et al. (2007) i.e. averaging crop calendars over regions of strong climate and topographical gradients would likely lead to unrealistic dates.

It is noted however that in practice and at fine resolutions and farm level spatial extents planting dates are not only functions of climate (e.g. onset of seasonal rainfall) but also depends on non-climatic factors such as labour, availability of farm inputs, and other socio-economic factors (Deryng et al., 2011; Hansen et al., 2011; Jones et al., 2000; Wolf et al., 2015). The derived dates however compare well to the FEWSNET calendar already described in Section 5.3.

5.4.2 Harvest dates

Harvest dates are influenced by crop growth and maturity that are constrained by the bio-physical nonlinear interactions between climate, soils, cultivars and management practices. As such, harvest dates determined by these interactions resulted in dates that were quite non-uniform even in regions with homogeneous rainfall regimes. The length of a growing period is determined by thermal time accumulation i.e. effective temperature from emergence to anthesis (TSUM1) and affective temperature from anthesis to maturity (TSUM2) in WOFOST. This results in different harvest dates even in region of equal rainfall amounts. The influence of thermal time accumulation is also highlighted in (Deryng et al., 2011). The differences between forecasted harvest dates and the reference (i.e. less than 10 days) however illustrate good simulation of the bio-physical interactions in WOFOST.

5.4.3 Deterministic validation

Simulated WFDEI yields show differences compared to FAO and NAT yield statistics. The difference may be due to a variety of reasons. One reason could be the single cropping season in the current WOFOST setup and non-climate related factors that influence observed yields (such as availability of seeds, fertilizers, change in technologies or civil strife amongst others) but are not represented in WOFOST. All countries in the study area have two seasons of rainfall, and hence cropping. Configuring WOFOST to simulate double cropping as opposed to the current setup may improve simulated yields in comparison to the observed. Methods used to record observed statistics may also cause differences, for example, FAO only includes the bulk of harvests in a particular year based on sample agricultural production surveys. Any harvests in the latter part of the year are added to the following year's statistics. In WFDEI yields we aggregate all harvests within one year.

Differences between FAO and NAT yields (and accumulated sub-national yields; not shown) highlight the uncertainty in Observed yield statistics and yield statistics collection methodologies.

Existing differences between WFDEI and reforecasted yields could result from a variety of factors. Since sowing dates, maize varieties, and fertilizer applications rates are fixed in both the WFDEI and reforecasted yield simulations, any differences must be attributed to variations in the climate input fields. Changes in weather distribution and characteristics not only modify soil characteristics, but also interact with maize physiological processes during the growing season. For example, because it is not possible to provide the exact distribution of weather during a season from climate forecasts, the distribution of dry-wet day frequencies and temperatures below or above critical thresholds would each individually, or acting in combination grossly influence carbon assimilation and partitioning to different organs. Even though we corrected biases in climate forecasts against the reference climate in a bid to capture their daily distributions, any chronological differences would still trigger nonlinear interactions with maize physiological processes. All these complex non-linear interactions result in simulated yields different from the observed and are probably responsible for the difference between forecasted and WFDEI yields. An explanation of some of the weather-climate related interactions are found in Ceglar and Kajfež-Bogataj (2012) and Porter and Semenov (2005). Existing difference may also result from fixed sowing dates, for example, assuming that rainfall is the main determinant of sowing dates and emergence, fixed dates in model simulation may result in non-emergence in some years and seasons or poor timing of crop phenological stages and hence affecting harvests. Improvements on sowing dates can be made by allowing evolving weather to determine sowing and emergence dates. Detailed evaluation at a single grid point or smaller spatial extent may highlight the causes of complete losses in some planted areas. Still, good/poor skill at a single grid point or land use cell cannot be generalized over a larger area if neighboring cells do not show skill.

In seasonal prediction, whether yields or climate, anomalies are of more interest than the forecast of absolute values. In the following paragraphs therefore, we present a discussion on probabilistic forecasts of anomalous below-normal and above-normal yields.

5.4.4 Probabilistic prediction skill

We employ an ensemble based yield prediction to assess the skill of probabilistic yield forecasting. The use of historical WFDEI climate data to simulate reference yields assumes perfect knowledge of weather during the growing season. Ensemble yield forecasting enables us to include uncertainty information in yield forecasts even though the current study did not investigate sources of uncertainty in the modelling chain (i.e. from climate to yield forecasts). Any improvement in forecast skill above climatological forecast therefore illustrates the value gained by ensemble forecasting. We have used RPSS and ROCSS to evaluate the overall ensemble yield forecast skill

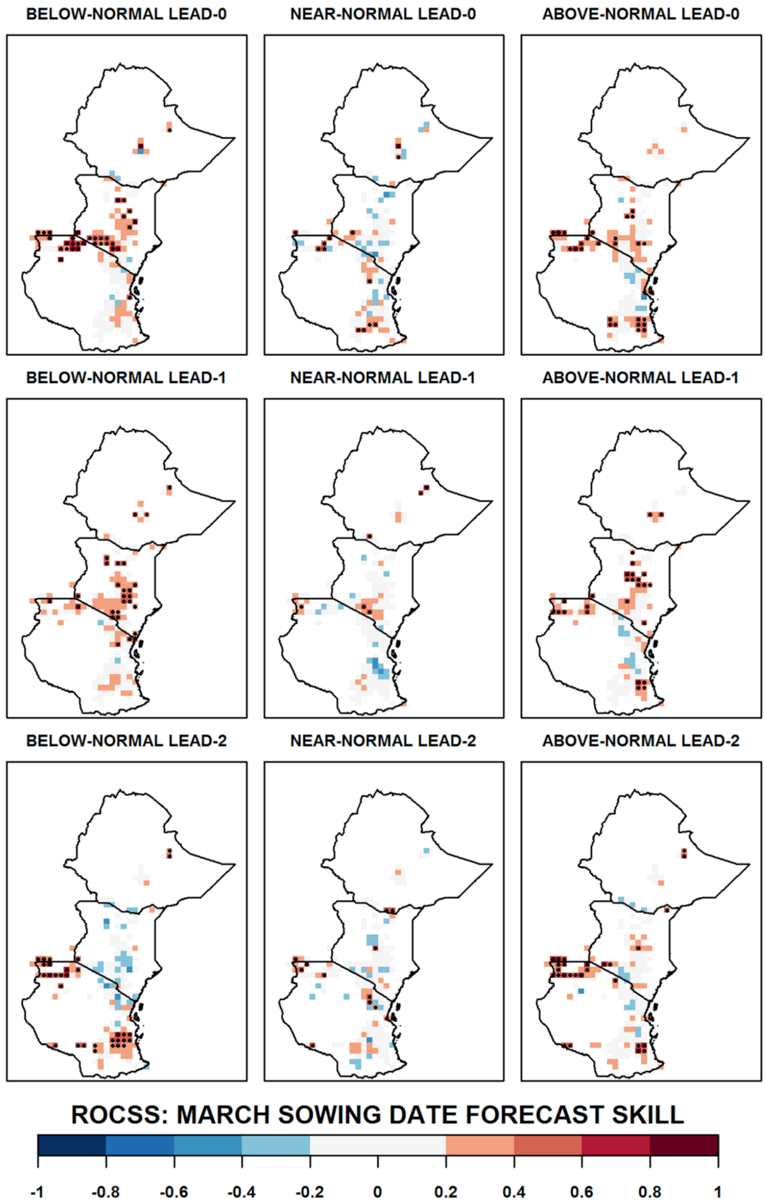


Figure 5.9: Grid point S4 yield ROCSS for March sowing date and 0-2 lead-months. Red colours show regions of skill while blue colours are regions with no skill. ROCSS of -0.2 to +0.2 show regions of no forecast improvement over a climatological forecast. Dots indicate areas where ROCSS is significantly positive at 90% confidence level.

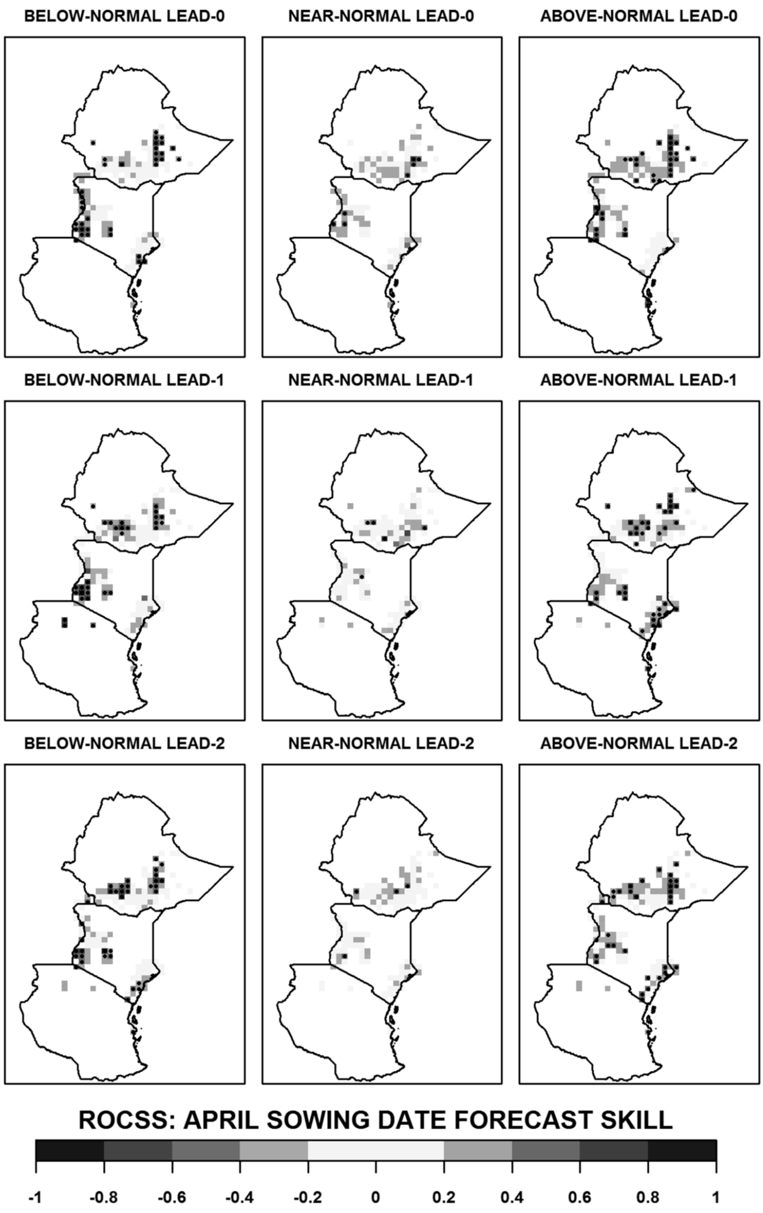


Figure 5.10: Same as Figure 5.8 but for April sowing date.

and ROCSS to evaluate performance in predicting anomalous yields (i.e. below, near or above tercile yield thresholds). Even though the use of RPSS and ROCSS is established in atmospheric sciences to evaluate ensemble climate forecasts, the metric and its results can also be meaningfully applied to other ensemble climate impact forecasts, such as hydro-logical forecasts or crop forecasts like in our study. ROCSS has been applied to assess skill of ensemble seasonal forecasts of impacts such as yield prediction (Cantelaube and Terres, 2005; Challinor et al., 2005; Marletto et al., 2005), hydrological forecasts, and malaria prediction (Marletto et al., 2005; Morse et al., 2005) among others.

5.4.4.1 Ranked probability skill score (RPSS)

RPSS evaluates closeness of cumulative probabilities of forecasts and observation vectors relative to a climatological forecast. This measure is sensitive to the number of ensemble members. It results in negative bias for fewer members (Kumar et al., 2001; Mason, 2004) and increases positively when more ensembles are used. Buizza and Palmer (1998) assessed the impact of ensemble size on ECMWF seasonal cli-mate forecast (varied ensemble members from 2 to 32 in multiples of 2) and concluded that increasing ensemble size improves ensemble prediction skill but the effect is also dependent on the specific performance metric. Kumar et al. (2001) conclude that skill achieved by 10–20 members in climate predictions is sufficient and close to the expected average skill from infinite ensemble size. Using fifteen (15) ensembles in this study should be sufficient to assess yield prediction skill. The score is increasingly penalized the more its cumulative probabilities deviate from the reference cumulative probabilities (Weigel et al., 2008, 2007a,b). RPSS values obtained in this study are significantly positive in only a few grid points. This we attribute to the fact that RPSS does not consider skill of individual probability threshold forecasts but rather the cumulative distribution of the forecast. Even though good skill may be achieved from a single tercile probability threshold and poor skill in others, the final score would be low or even negative when averaged over all tercile thresholds. Negative or low values of RPSS has however been noted to hide useful information regarding forecasts (Mason, 2004). This may be true for this study.

Assuming we have sufficient ensemble members, we conclude therefore that low and negative RPSS scores result from the distance between the cumulative yield forecast probability curve and the reference yields but not from ensemble size. However a study involving more ensemble can give a definite conclusion on how the size influences RPSS in crop prediction.

5.4.4.2 Anomalous yield prediction

We use the ROCSS, a skill measure with values ranging from -1 (perfectly wrong forecast) through zero (no skill forecast) to 1 (perfect forecast). We assessed forecasts of above-normal (upper tercile), below-normal (lower tercile) and near-normal (middle tercile) yield anomalies.

There is significant positive prediction skill of anomalous low (crop failure) and high yields. This is true for at least 15–35% of the areas planted with maize in a par-

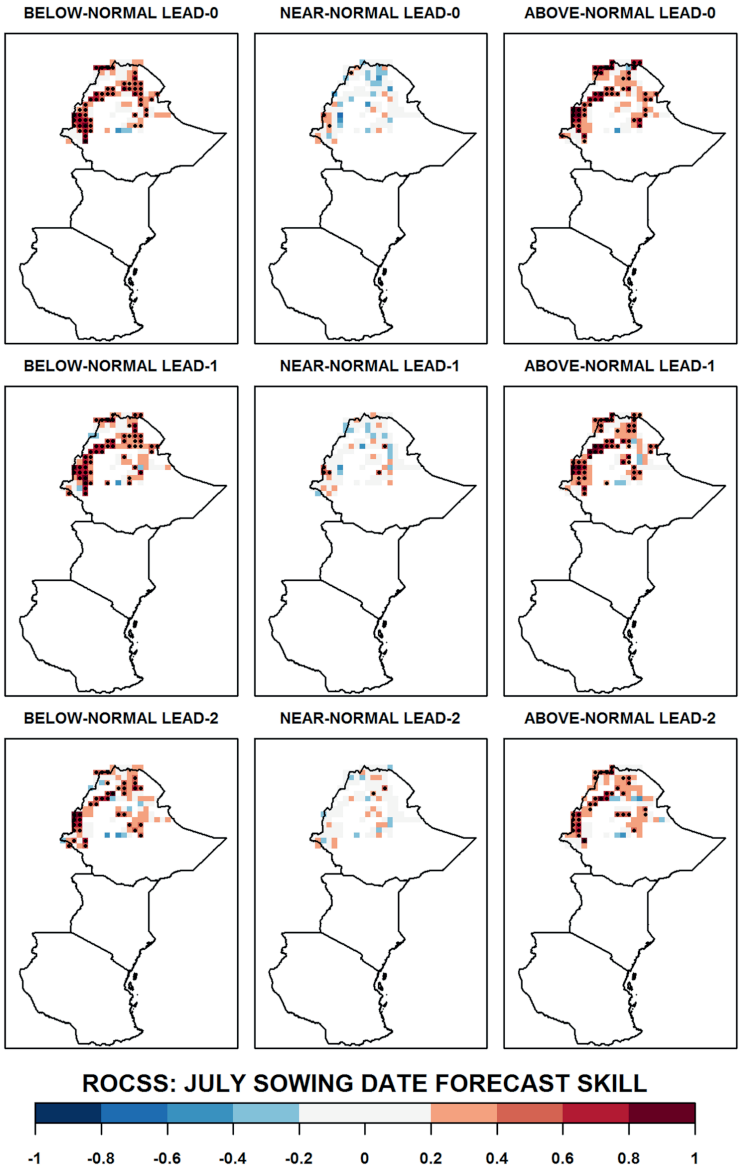


Figure 5.11: Same as Figure 5.8 but for July sowing date.

ticular season. The fact that areas of significant skill are continuous, i.e. consist of multiple adjacent cells that are significant, and persistent, i.e. extends over multiple lead months, lends even more credibility to the potential value of seasonal crop yield prediction. Areas worth mentioning in this context are for March sowing in Tanza-

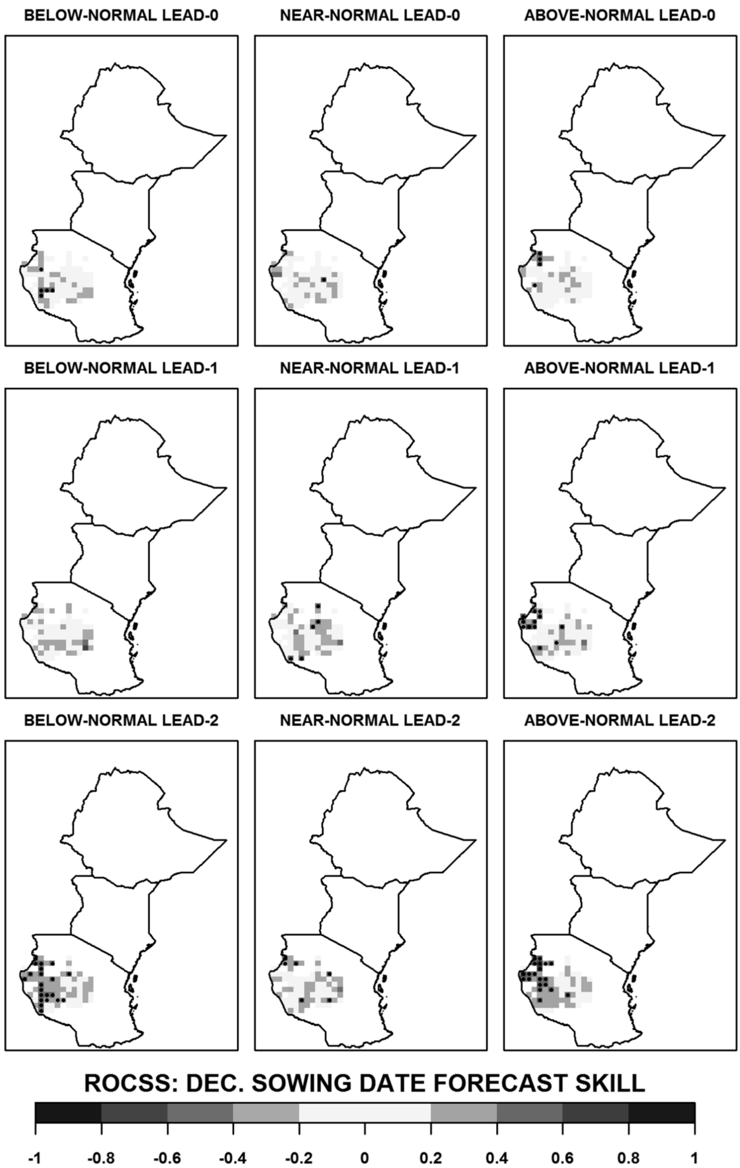


Figure 5.12: Same as Figure 5.8 but for December sowing date.

nia: Kagera, Mwanza, Mara and Lindi. For March and April sowing in Kenya we found skill in its south western regions (Western, Nyanza, Narok, Kajiado), its Coast and some areas in central Kenya (Laikipia, Meru). For July sowing in Ethiopia we have skill in large parts of Tigra, Amhara, Benishangul-Gumaz and western Oromia

(Shewa, Wellega). Even though good skill of at least 2-lead months are achieved in this study, prediction over crop growing period can be improved by for example updating yield predictions based on updated climate forecasts during crop growing season or by improving physical processes in WOFOST for example improved soil moisture maps or including a soil sub-model that can account for detailed processes plus double cropping where it exists.

Such information would greatly benefit farmers, governments, policy makers, relief agencies, agricultural commodity traders and even crop insurance companies if done at suitable spatial extents.

5.4.5 Implications of the study

Seasonal climate forecasts are operationally issued to users accompanied by information on expected impacts on various sectors, including agriculture, water resources, etc. Over East Africa and the greater horn of Africa (GHA), consensus seasonal climate forecasts are issued by the greater horn of Africa regional climate outlook forum (GHACOF) as the probability of exceedance of a certain rainfall threshold, or of rainfall being within a range of thresholds (Husak et al., 2011). This information must be understandable to the recipients if expected to influence any change in farm management activities, policy decisions either by relief agencies or governments to better manage associated climate risks (Troccoli et al., 2008; Jones et al., 2000). Forecasts should contain related uncertainty information. According to Jones et al. (2000), uncertainty limits the benefits of seasonal climate forecasts but Troccoli et al. (2008) notes that climate information and predictions play an essential role in management of risks associated with climate variability and change.

A number of early warning systems and organizations (described in Section 5.1) exist in East Africa. These organizations provide their services largely to governments and humanitarian agencies and do not target the local farmers. For crop monitoring they use agrometeorological assessment reports and satellite technologies that monitor conditions of food crops after planting (e.g. NDVI) and rainfall to estimate impending food security situations. FEWSNET for example uses seasonal climate forecast in its assessments but does not provide explicit yield forecasts. This study can feed into the existing EWSs by providing pre-season yield forecasts to farmers through farm extension officers based on yield forecasts in their locations (because they can adjust their farm management decisions), farm input commodity traders (for stocking of farm merchandise) besides extending the time horizons of the existing EWSs by providing forecasts of expected yields before planting.

Even though the results shown in this study are aggregated to half degree grid (weighted by planted area) and further to national country boundaries, the half degree provides a resolution presently lacking and may be useful to governments, insurance, and relief agencies because it captures the sub-national variability. The WOFOST crop model simulation in this study is done at FAO land use cells, an even finer resolution product that may be useful for farmers. We plan further work on this finer resolution FAO land use cells. The Climate Change Agriculture and Food Security (CCAFS) propose to develop an Integrated Agricultural Production and Food Security Forecasting System (INAPES) for East Africa (Kadi et al., 2011). This system will integrate improved seasonal climate forecasts, yield and pricing forecasts at high temporal and spatial resolution to enhance the early warning systems. With such developments, this study is timely as it may provide useful insights. We have assessed seasonal climate skill for East Africa (Ogutu et al., 2017), and have gone ahead to demonstrate its usability in crop yield fore-casting in this paper. Our study can directly feed into the development of new systems envisaged for the future such as INAPES.

Cited references in this manuscript have evaluated yield prediction skill focusing on several months lead time before harvesting. In this study we evaluate yield prediction prior to sowing. To the best of our knowledge, this is a first attempt to perform such pre-season evaluation using dynamic GCM climate forecast and a crop simulation model over East Africa.

5.5 Conclusions

We have shown the potential of ensemble seasonal climate forecasts in combination with a crop simulation model to predict pre-season water-limited maize yields. The study showed that there is potential to predict maize yields for at least 2-months before planting dates. Yield forecast skill varies with regions and hence sowing/harvest date combinations, and lead time. Such lead time has potential to influence regional policy and local management decisions related to maize production, such as adjustments of farm management for the upcoming seasons.

While accurate crop calendar data is an essential input into crop models, there is lack of a high resolution regional data that can be applied in regional studies. The calendars used are sufficient to assess maize predictions even though we recommend a detailed study to assess the sensitivity of ensemble yield predictions to planting dates through a case study or field experiments. Incorporating the use of remote sensing to derive dates such as emergence could play a role in improving the planting dates.

To a large extent, the spatial patterns of mean yields and harvested grid cells are well simulated. Interannual variability in forecasted yields are higher than the reference in most of the grid cells but depend on the growing seasons (i.e. sowing dates).

In general, distribution of grids with significant RPSS varies with sowing dates/harvest dates and regions of the study area. The RPSS generally results in low skill as opposed to the ROC and ROCSS that has shown good skill for anomalous low and high yields. This emphasizes the importance of using more verification measures since they quantify different attributes of a forecast.

Since the aim of seasonal forecasts, whether climate or its impacts, is in prediction of anomalies, there is in Eastern Africa a definite and significant potential to provide seasonal crop yields prediction based on seasonal climate forecasts. Predictions may be improved further if the crop simulation model was configured to simulate double cropping in regions where it occurs.

Acknowledgment This study was financially supported by the EUPORIAS project (European Provision of Regional Impact Assessment on Seasonal-to-decadal timescale); grant agreement No. 308291 funded by the European Commission in the Seventh Framework Programme for Research. We thank the anonymous reviewers whose valuable comments helped improve this article.

Chapter 6

Influence of spatial aggregation from maize yield simulation grid on predictability: a case study

Abstract

The purpose of this study is to examine the influence of crop model output aggregation on sub-national crop simulation and predictability of seasonal climate driven maize production forecasts over East Africa. We assessed this using two case study regions (i.e. North-Gonder in northern Ethiopia, and Bungoma in the equatorial region of Kenya). The WOrld FOod Studies crop simulation model (WOFOST), a field scale crop model is configured to simulate maize production over East Africa at a high spatial resolution of 0.1° grid corresponding to input soil data grid. Rainfall, temperature and downward shortwave radiation seasonal climate forecasts for the period 1981-2010 from the European Centre for Medium range Weather forecasts are fed into WOFOST. Climate grids are coarser (0.5°) implying that several 0.1° grid boxes are within climate grids and thus have similar climate characteristics. This likely resulted in crop simulation missing the detailed small scale crop-soil-climate interactions especially so in a region with steep rainfall variability such as East Africa. We assessed the influence of aggregation over sub-national boundaries by soil types and by crop varieties using the two case study areas. Influence of aggregation on average yields depend on physiographic characteristics and area of sub-national administrative units. After aggregation to sub-national regions, higher predictability is exhibited in larger areas compared to the smaller ones, perhaps influenced by the number of climate grid

This chapter is under review for publication as:
Ogutu G.E.,I. Supit,P. Omondi and R.W. Hutjes, 2020: Spatial Aggregation Units' Influence on Predictability of Seasonal Forecasts of Maize Production: a case study. Agricultural Systems Journal

cells over each region. Crop production simulation and predictability is similar over the dominant soil types within the case study areas (i.e. loamy sand-to-sandy loams; and sandy loams-to-loam soils). This may be because of similar available soil water characteristics. Good, significant prediction skill in individual high resolution grid cells are degraded little when averaged over sub-national boundaries. The higher the number of climate grids in a spatial extent, the better the crop simulation and predictability implying that higher resolution climate data may improve simulation and predictability. The results of this study i.e. skill assessment at national boundaries and high resolution crop simulation units may inform both maize production related policy decisions at regional or national levels, and also support maize production decisions at specific cropping locations such as farm management decisions made by farmers.

6.1 Introduction

Potential benefits can be extracted from the use of seasonal climate forecasts for agricultural production. Also, significant progress has been made in modeling agricultural and climatic processes and some of their interactions (Baron et al., 2005) but a challenge of linking climate forecast to agricultural impact models exists due to the different spatial and temporal scales. Such challenges include deriving appropriate input information from climate forecast models, mainly Global Climate Models (GCMs) for crop models, validation by ground data that is at times unavailable or of poor quality, and the effect of scale on the laws governing processes relevant to crop growth and development. Even though climate-yield relationships are quite complex, robust relationships can be established between regional atmospheric circulations, surface climate, and crop productivity . The patterns of seasonal climate and their impacts are important for understanding vulnerability and adaptation of regional agricultural production. Zhao et al. (2015) notes that even though there have been increased development of agricultural systems, rain fed crop production is still grossly influenced by meteorological variables. Precipitation remains a determinant in many regions of the world and water deficit is one of the most significant factors of crop production. Therefore, understanding climate-crop relationships and developing tools to predict crop production would help develop risk management strategies. But the forecasts are important if used appropriately with an understanding of their capabilities and limitations (Hansen and Ines, 2005; Skees et al., 1999).

This makes communication of associated uncertainties to facilitate apportionment of confidence levels to the predictions necessary. Targeted users of climate forecasts and its impacts will only act if they have confidence in the products, and this applies to all users irrespective of their level of understanding of forecasts. For example farmers, governments, aid agencies whether local or international, insurance companies and even disaster risk reduction practitioners would require tailored forecast information in different formats.

A climate information service should then deliver understandable impact predictions to stakeholders considering regions and periods of high or low skill and uncertainty in prediction.

Because crop models are basically designed for farm level studies, implementing their use in regional studies requires spatial aggregation of the yield or biomass output, weighted by the fraction of areas cultivated or harvested within regional boundaries. Alternatively, regional crop production simulations may be conducted at particular sites or grid-boxes across smaller or larger areas but with soil and climate characteristics spatially averaged over the simulation units. Yet again, input data aggregation have been found to influence simulations of yields (Folberth et al., 2012; Supit and Van der Goot, 1999; Olesen et al., 2000; Porwollik et al., 2017), phenology (Van Bussel et al., 2011), and biomass (Kuhnert et al., 2017). Country masks used for aggregation of simulation outputs has been found to influence simulated yields due to variability in harvested areas (Porwollik et al., 2017). Input data aggregation does not however wholly capture the interactions between crop varieties, soil types and climate characteristics at fine resolutions for which crop models were originally designed. Because of highly variable local climate-soil-crop interactions, regional forecast variations do not necessarily represent local variations representative of a farm scale (Martin et al., 2000). For example, local weather variation can lead to dry spells at higher spatio-temporal resolutions than is captured when the outputs are aggregated over a region. As such, farm scale variability has been noted to be higher than regional variability (Martin et al., 2000).

Applied in yield forecasting, aggregation introduces uncertainties in the forecasts and may either cause reduced prediction skill, or may average out simulation errors thus resulting in better prediction. Even so, spatial aggregation of crop production suffices for sub-national, national or regional crop performance assessment (Olesen et al., 2000), and is sufficient to inform policy decisions mostly operationally made at regional units (Van Bussel et al., 2011; Folberth et al., 2012). Comparing the variability of yields aggregated over a spatial region to actual observed yields (official yield statistics); largely collected via agricultural census would give an indication of the influence of aggregation in the current study.

Prediction results at regional scales are useful mainly to governments, relief agencies, and regional early warning organizations. Farmers need high resolution information near farm-scale that would enable adaptation to the expected climate over a crop growing season . In the Greater Horn of Africa (GHA) sub-region, pre-season crop yield forecasting driven by seasonal climate forecast has been found skilful for at least two lead months prior to sowing in some regions and seasons (Ogutu et al., 2018) based on crop production data aggregated to half degree resolution from a high resolution simulation unit of 0.1-degree. A number of studies on crop prediction have been carried at 0.5° being close to the nominal resolution of many gridded input climate data. This resolution has been suggested as useful for regional scale analysis (de Wit et al., 2010) and has been used for crop prediction studies in many regions of the world . In

Ogutu et al. (2018), half degree climate input data with high resolution 0.1-degree soil characteristics and land use data (soil mapping units, SMU) was used with optimised crop varieties in crop simulation units as well. This means that several soil resolution grids falling within single climate half degree grids had similar climate characteristics and variability. Even though this method may not capture the highly variable climate (i.e. rainfall) in the tropical regions, forecasting at 0.1-degree resolution (near farmers' field-scale) without aggregation to neither half-degree nor county boundaries could epitomize the usefulness of crop forecasts to farmers. It is therefore necessary to assess crop production forecasts at such high resolution before upward aggregation.

Among different soil types and crop varieties, yields may vary due to differences in nutrient availability, soil nutrient retention capacity and water holding capacity and the interactions between crop characteristics and soil types for example, the maximum root depth and plant water available capacity are assumed to vary between soil types. Simulated yields at field scales have been shown to respond differently to climate variations on different soil types which would make the response of aggregated yields strongly dependent on the soil-climate-cultivar interactions. Soils determine the crops' suitability and offer a platform for cultivars, fertilizer and moisture interactions. Aggregation camouflages these small scale interactions, thus making estimation of the optimum scales for estimating (e.g. county and national) yields difficult. Aggregation can cause large systematic errors due to non-linear crop behavior and may lead to scale mismatches . Evaluating the effect of aggregation on prediction skill over varying soil types is therefore fundamental in order to make forecasts useful.

Similarly, crop cultivars show varying sensitivity to soil types, fertility, and climate characteristics at different SMUs. The sensitivity to temperature and rainfall for example are cultivar specific influenced by many processes such as but not limited to evaporation and transpiration. The variability in soils and climate may affect crop processes in one cultivar variety but not necessarily the same processes in another cultivar (Porter and Semenov, 2005) resulting in good performance of one cultivar in a large spatial area or SMU but performing poorly in another. Evaluating the performance of different cultivars in varying soil types, geographical regions, plus the influence of aggregation on each variety over sub-national spatial regions help in understanding crop production predictions.

This study therefore seeks to understand the influence of crop production aggregation over a sub-national region referred to as "Nomenclatures des Unitẽs Territoriales Statistiques level 2 (NUTS2)" in (Supit and Van der Goot (1999) and Boogaard et al. (2013)) on seasonal maize forecasts by:

1. assessing the influence of yield aggregation on predictability by comparing the official observed agricultural statistics to simulated NUTS2 crop yields aggregated upwards from both half-degree and 0.1-degree resolutions

2. analyzing and comparing influence of aggregation on predictability based on 0.5° and 0.1° resolutions

3. assessing predictability by aggregation over similar soils and varieties.

In all cases we want to assess how the original grid before aggregation affects predictability of both yields and biomass (TAGP).

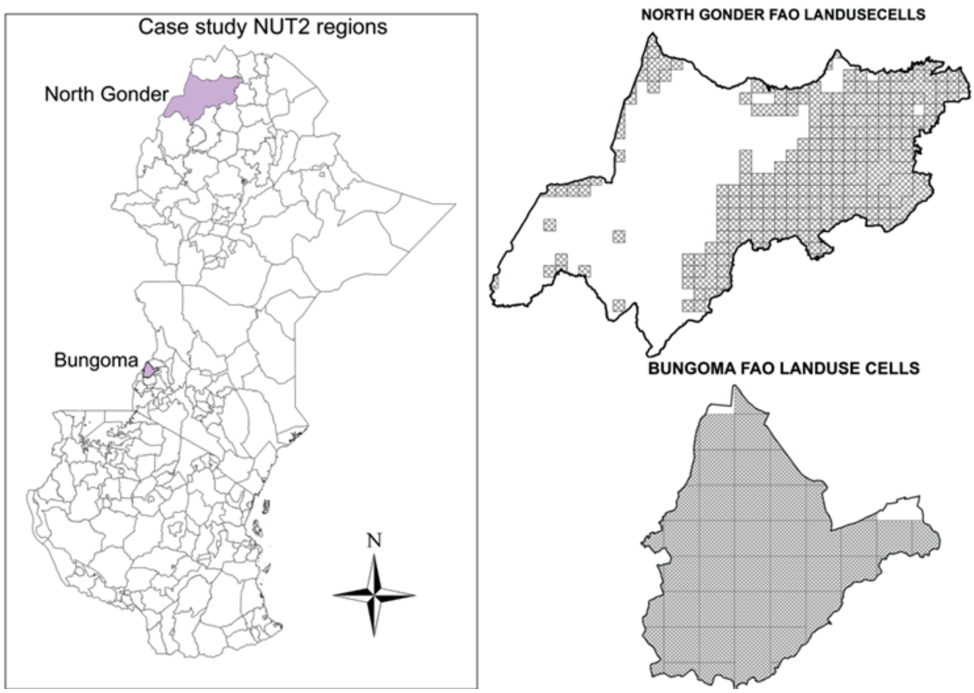


Figure 6.1: Map of study area, case study NUT2 regions and FAO land use grids (approximately 0.1° resolution)

6.2 Methodology

6.2.1 Model description

We simulate hindcast maize production for the period 1981-2010 using the WOrld FOod STudies (WOFOST) crop simulation model, a detailed model with respect to crop physiology allowing for specification of regionally used crop varieties. It was originally developed to simulate crop yield for a single location with homogeneous cultivars, soil characteristics and different weather. The model is photosynthesis driven and simulates daily growth and production of annual crops using a range of physiological processes from sowing to maturity in response to weather, soil type, and soil moisture conditions as defined by crop characteristics during growth season.

The physical processes comprise light interception, photosynthesis, respiration, evapotranspiration, assimilate partitioning, leaf area dynamics, phenological development, and root growth (Supit et al., 2010; Boogaard et al., 2013).

In this study, the model simulates theoretical water-limited-yields (WLY) at simulation units of 0.1°×0.1° resolution determined by overlaying land use, elevation, crop calendar, soil maps and climate data. All overlapping features in the input layers are clipped and extracted to create soil mapping units (SMU) or simulation units. For WLY simulation, WOFOST soil profile is divided into two layers; the upper rooted zone and the lower zone between actual rooting depth and maximum rooting depth. The first zone and second zone merge gradually as the roots grow deeper. The groundwater is so deep that it does not influence the soil water content in the rooting zone. WLY is influenced by rainfall assuming a soil nitrogen level that reflects a maximum yield level of about 40-50% of the potential, typical for the region, and otherwise optimal crop management and that no losses occur due to pests or diseases. In the current setup, the first cropping season in a year is modelled from emergence to maturity determined by the thermal time required to reach the development stages (TSUMS). See Chapter 2 for further description of the model.

Table 6.1: Maize varieties used in this study as defined by WOFOST thermal time parameters. *Note*:TSUM1 = temperature sum from emergence to anthesis (°C day); TSUM2 = temperature sum from anthesis to maturity (°C day); both relative to a base temperature of 10°C

MAIZE VARIETY	TSUM	VALUE	REGION
1	TSUM1	400	North-Gonder
	TSUM2	350	
3	TSUM1	620	North-Gonder and Bungoma
	TSUM2	570	
5	TSUM1	720	North-Gonder
	TSUM2	670	
6	TSUM1	770	North-Gonder and Bungoma
	TSUM2	720	
7	TSUM1	820	North-Gonder
	TSUM2	770	
9	TSUM1	920	North-Gonder
	TSUM2	870	
10	TSUM1	970	North-Gonder
	TSUM2	920	

6.3 WOFOST input data

6.3.1 Soil and land use data

Land use and soil input data are derived from the 0.1°×0.1° grid International Soil Reference and Information Centre-World Soil Information (ISRIC-WISE) database (and http://www.isric.org/data/isric-wise-international-soil-profile-dataset). The database includes information on soil physical characteristics, root depth, and landscape characteristics such as elevation, slope gradients, and slope aspects. Soil properties such as wilting point and field capacity are estimated with the pedo-transfer functions from Saxton et al. (1986). Maize growing areas are determined based on FAO land use maps . The intersection between ISRIC-WISE soil data base, FAO land use maps and country administrative region maps (known as NUTS regions) and weather data results in SMU, the basic WOFOST 0.1°×0.1° simulation unit in this study. The SMUs are assumed to be spatially homogeneous with respect to planted crop varieties, weather and soil characteristics. NUTS structure is such that the highest level follows the national country boundaries; NUTS-0 which is further divided into regions (NUTS-1) and NUTS-1 are divided to sub-regions (NUTS-2). The method used to formulate modelling units is congruent to methods by, among others, Supit and Van der Goot (1999) and Resop et al. (2012).

6.3.1.1 Weather data

We use weather data (i.e. precipitation (*tp*), maximum (*tmax*) and minimum (*tmin*) temperatures, surface downward shortwave radiation (*rsds*) from a re-anlysis (ERA-Interim) and observation fused data product from the Water and Global Change (WATCH) forcing data ERA-Interim (WFDEI) at 0.5°×0.5° resolution (Harding et al., 2011; Weedon et al., 2010, 2011, 2014) to simulate reference yields. WFDEI is obtained after elevation and monthly bias corrections of the ERA-Interim re-analysis (Dee et al., 2011) using gridded observed climate data from University of East Anglia Climate Research Unit (Harris et al., 2014). ERA-Interim is one of several available climate reanalysis products providing a numerical description of the recent climate produced by assimilation of weather observations in forecast model simulations and are considered the most consistent multivariate representation of the past climate (Dee et al., 2011). We use WFDEI to represent the observed climate for the period 1981-2010.

To simulate yield forecasts, we use climate reforecasts (hindcasts) from the ECMWF System-4 seasonal climate ensemble prediction system , bias corrected at grid level for each initialisation and target month combination separately against WFDEI. Seasonal climate forecasts are initialized on the first day of every month from 1981-2010 with 15 perturbed (different) initial conditions. Since weather evolution during an upcoming season is not certain, initializing the climate forecast model from different initial conditions allows sampling of the possible weather evolution path during the coming season. Forecasts initialized from each of the perturbed initial conditions results in 15 different forecasts (ensemble members). Each member provides a forecast 7-months into the future. We will only consider forecasts starting in the month of sowing (lead-0) or one or two months prior to sowing (lead-1 and lead-2).

Predictability of the seasonal climate forecasts over East Africa is analysed in Ogutu et al. (2017) in which useful probabilistic prediction skill is found in all major cropping regions of East Africa, but the skill level decreases from near surface air temperature; to precipitation; to downward surface radiation. To simulate yield forecasts, WOFOST is driven with each of the 15 climate forecast ensemble members; resulting in 15 ensemble yield forecasts, i.e. it spans the range of possible yields resulting from the 15 probable climate evolution during the crop growing period. This enables an assessment of probabilities of anomalous forecasts.

6.3.1.2 Crop data

In this study, we have used a method that combines the optimization of sowing dates on maximum yields; and optimization on climate and crop specific characteristics to determine maize sowing dates at 0.1-degree resolutions. Details of these methods are described in chapter 2 and Ogutu et al. (2018). The procedure starts with coarse crop calendars from Sacks et al. (2010), and standard tropical maize varieties for crop modeling compiled in Van Heemst (1988). Yields are simulated for the different maize varieties planted at every grid point every 10 days, starting 90 days before and end 90 days after the sowing dates presented in the coarse crop calendars, using the baseline climate forcing (WFDEI) for the period 1981 to 2010. Maize varieties differ mainly in terms of the thermal time that is needed to reach flowering and maturity calculated based on the planting dates and a base temperature of 10°C. We subsequently selected the sowing date and crop variety that provided the highest average crop yield over the baseline period, resulting in 10 varieties in East Africa (see Table 6.1. The derived sowing date for each grid point was fixed (i.e. not varying between years) in WOFOST, while harvest dates are determined by the weather conditions during growth period.

6.3.2 Choice of case studies

For our case studies we select two sub-national regions (NUT2), North-Gonder (in Ethiopia) and Bungoma (Kenya) NUT2 regions shown in Figure 6.1. Choice was based on the following criteria; regions of good maize yield predictions based on findings in Ogutu et al. (2018); skillful climate prediction (Ogutu et al., 2017); areas of good records of actual observed (official) yields; and areas dominated by rainfed agriculture. North-Gonder lies between latitudes 11.75–13.75N and longitudes 32.25–13.75N, with surface elevation ranging from 527-4000 metres above mean sea level (masl). Bungoma spans longitudes 34.3–35.1E and latitudes 0.43–1.15N, with elevation of 1249–4000 masl. It would be useful to know if the same level of predictability observed at 0.5 degree resolution is affected by scale of spatial yield aggregation.

We have 25-years and 10-years of observed yield data for Bungoma and North-Gonder respectively, enabling a comparison with model simulated average yields, though not long enough to calculate more probabilistic skill metrics.

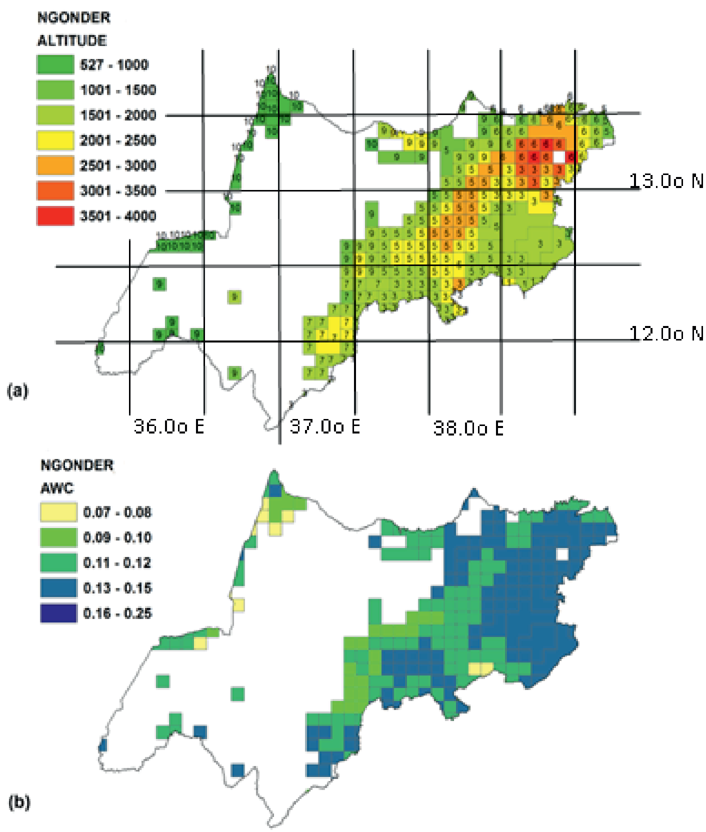


Figure 6.2: North-Gonder altitude ((a) in colours); crop varieties ((a) in numbers) and soil types shown in terms of available water capacity (AWC) in figure (b). Fine grid lines delineate SMUs; bold grid lines the 0.5° cells; white cells indicate non-agricultural regions.

North-Gonder has a complex terrain with altitude range of 527m to 4000m amsl and soil available water capacity range of 0.07 to 0.25% i.e. sandy soils in low latitude regions to silt-loam soils in the high altitude areas (see illustration in Figure 6.2). Seven crop varieties differing in their TSUMs (see table 6.1) are used in North-Gonder. The complexity in terrain, soils and crop varieties make this region a good candidate to assess complexities related to the influence of spatial aggregation on forecast skill. Predictability in this region is contrasted to predictability in Bungoma (Kenya), a region of less variability in altitude (though a high terrain), few crop varieties (see Figure 6.3a) and less variability in soil types (see Figure 6.3b)

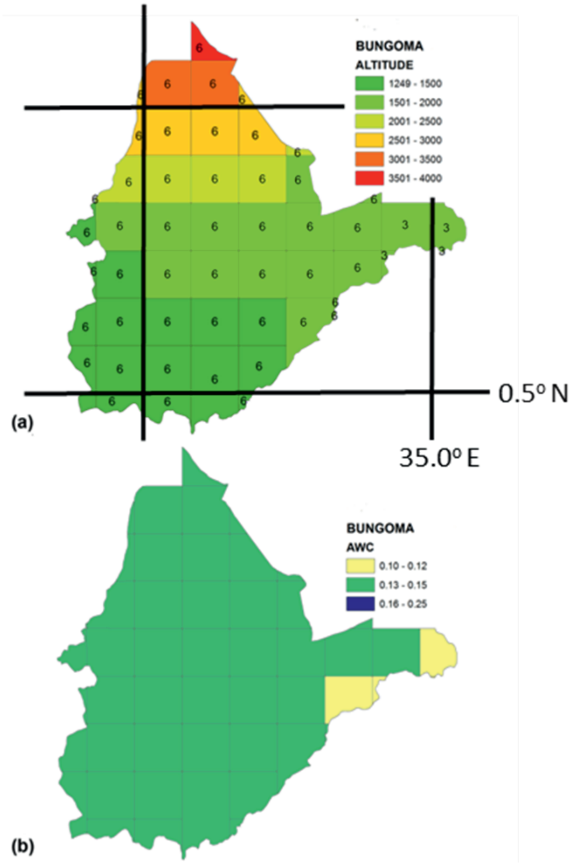


Figure 6.3: Bungoma altitude (figure (a) colors) crop varieties (Figure (a) numbers) and soil types shown in terms of available water capacity (AWC) in figure (b). Fine grid lines delineate SMUs ; bold grid lines the 0.5° cells

6.3.3 Assessment of crop production simulation and predictability

6.3.3.1 Model simulation

To assess skill of model simulation, good observed statistics are required. In this study we require good maize yield statistics at sub-national levels; in many instances such data are difficult to get. We use the mean and coefficient of variation (CV) to compare the degree of spread of yields, and error indices described in Willmott (1981); Willmott and Matsuura (2005) to compare the observed and simulated values. The metrics include mean error (bias), percentage bias (pbias), the root mean squared error (RMSE) and the index of agreement between the simulated and observed (d).

6.3.3.2 Predictability

We assess the effect of spatial aggregation on yield forecast skill using the relative operating curve skill score (ROCSS) to show skill of anomalous above normal (upper-tercile), below normal (lower-tercile) and near normal (middle-tercile) forecasts. For probabilistic forecasts, a warning is issued when the forecast probability for a pre-defined event exceeds some threshold. A set of hit-rates and false alarm rates are determined and plotted for these thresholds (we use decile forecast intervals) thus generating a ROC curve. The area under the ROC curve (AROC) is a measure for forecast performance. Because there is skill only when the hit rates are higher than the false-alarm rates, the ROC curve lies above the 45-degree line for a skillful forecast (AROC > 0.5), and below (AROC < 0.5) for a forecast that is no better than the climatology. AROC is transformed into a skill $ROCSS = 2AROC - 1$, where $-1.0 \leq ROCSS \leq 1.0$. A score of 1.0 indicates a perfect forecast system; a -1 indicates perfectly useless forecast system, and zero indicates no skill. For details see Appendix B.1. ROCSS has been used to assess predictability of seasonal climate forecasts for the study region (Ogutu et al., 2017) and for assessing skill of 0.5° grid yield forecast (Ogutu et al., 2018). In this article we assess forecast skill for each SMU, and aggregated to NUT2 boundary. Aggregation is accomplished with equation 4.1. We then aggregate SMU yields and biomass based on soil types (AWC) shown in Figures 6.2 and 6.3. We report the ROCSS and its significance plus the percentage of SMUs possessing significant prediction skill.

6.4 Results

6.4.1 Case study characteristics

Figure 6.4 and Figure 6.5 show the mean growing season climate for North-Gonder, and Bungoma NUT2-regions respectively for rainfall, maximum, and minimum temperature for 1981-2010. The two regions have different climate characteristics with North-Gonder having a highly variable rainfall (CV = 12%-30%), minimum temperature standard deviation of 0.4 to 0.5°C) and maximum temperature standard deviation (std) of 0.6 to 0.7°C). Variability in climate may be related to the variability in orography (Figure 6.2); and also the fact that North-Gonder has more weather grids (0.5°) than Bungoma. CV shown by North-Gonder rainfall emphasizes the complexity in simulations over regions of complex orography and climate. A combination of rainfall gradients, (i.e. range from 473 mm to almost 1000 mm) over the growing season combined with temperature and altitude gradients resulted in six maize varieties. Bungoma altitude and weather (from WFDEI) are not as variable (i.e. rainfall CV = 20.3%; maximum temperature std = 0.6 to 0.7°C; minimum temperature std = 0.6°C). Only two varieties are planted in Bungoma, with a single variety (variety-6) in over 90 percent of the simulation units (Figure 6.3a).

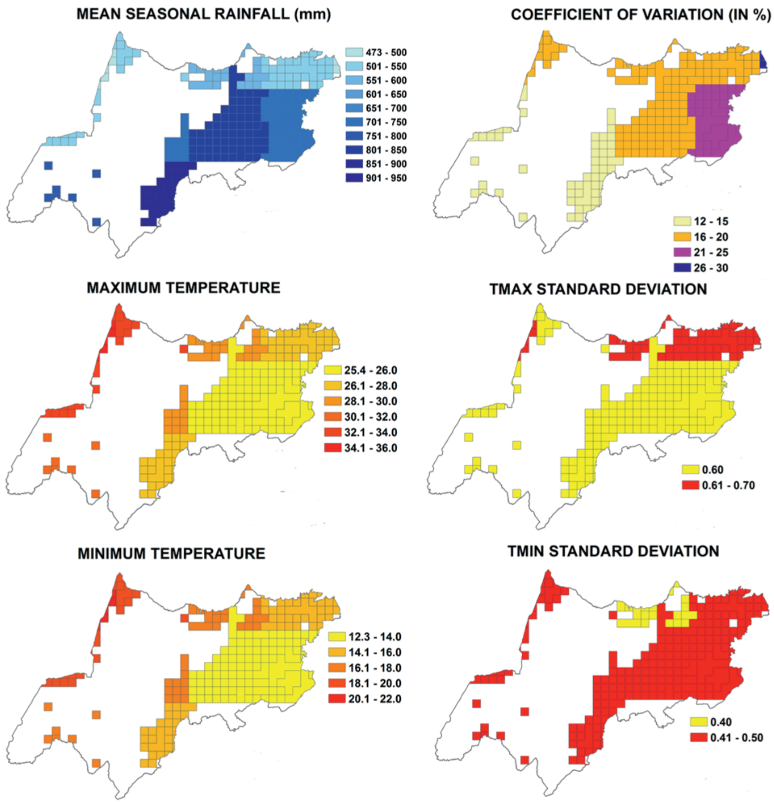


Figure 6.4: Mean cropping season climate of North-Gonder (JJAS) as represented by (a) seasonal rainfall (mm) (b) Coefficient of variation (%) (c) Maximum Temperature (°C) (d) Standard deviation for Tmax (e)minimum temperature and (f) Standard deviation for Tmin . Gridlines delineate SMUs; white cell are non-agricultural regions..

6.4.2 Influence of spatial aggregation on yield simulation and predictability

6.4.2.1 Comparison of model simulated yields to the observed

Table 6.2 show comparison between observed (OBS), WFDEI driven reference yields aggregated to NUT2 region from 0.5° resolution grids (NUT2_05), and yields aggregated from the SMUs (approximately 0.1°) grid. Observed yields are also compared to mean yield forecasts from 0.5° grid (05_FCST) and from 0.1° grid (FCST) for lead-time 0, 1, and 2. We show this for North-Gonder (Table 6.2a) and Bungoma (Table 6.2b).

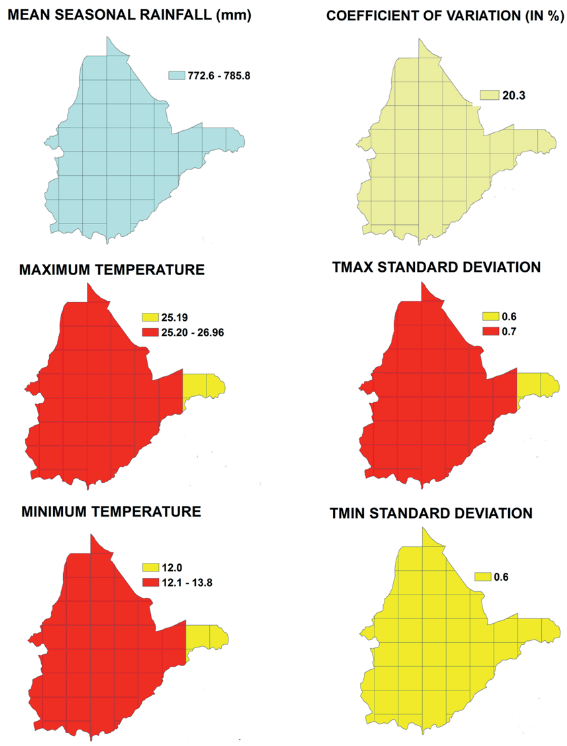


Figure 6.5: Same as figure 6.4 but for Bungoma (MAM season)

Half degree reference (NUT2_05) and forecasts (05FCST) at lead zero are generally not significantly different from the observed (OBS) in North-Gonder (Table 6.2a), however at 0.1-degree both NUT2 reference and forecast yields (FCST_LEAD0 to FCST_LEAD2) are generally higher and significantly different from OBS. In North-Gonder, higher CV is discerned in WFDEI reference yields aggregated from half degree grid than when aggregated from SMU (NUT2_05 = 9.0%, NUT2 = 7.3%). Higher CV in forecasts aggregated from half degree grid range from 11.4% - 9.8% while those from SMU are 8.1% - 6% for forecast lead-0 to 2 respectively. In this region, less spread is obtained from forecasts aggregated from SMU grid than from yield data aggregated from 0.5 degree grid, the latter decreases with lead time because perhaps cells that are not harvested are eliminated before aggregation resulting in reduction of aggregated cells with lead time. Hence reduced yields when weighted by cultivated area. This is as opposed to aggregation from 0.5° cell i.e. to generate yield data at 0.5-degree, simulated data is weighed by the percentage area cultivated within the grid box. These two approaches may result in the differences in spread.

In Bungoma (Table 6.2b), all simulated yields are lower and significantly different from OBS irrespective of initial resolution before aggregation. NUT2 yield aggregated from

SMUs has a higher CV equal to 22.9% compared to 19.2% from NUT2_05. Half degree forecasts 5FCST_LEAD0 and LEAD1= 11.3% and 14.6% while FCST_LEAD0 and LEAD1 = 18.1% and 15.4% respectively.

This may be because Bungoma is enclosed in a single 0.5° grid compared to North-Gonder, resulting in a relatively lower spread because of limited contribution from half degree grids. In both North-Gonder and Bungoma, observed deviation from the mean yield is higher than NUT2_05 and NUT2 deviations. Like the observed standard deviation, the percentage spread of data around OBS (CV) is larger than the simulated. However the percentages are region dependent with higher CV in Bungoma (both OBS and simulations) than in North-Gonder irrespective of aggregation level.

The effect of aggregation is that, extremes are evened out when aggregated from 0.5°resolution than when aggregated from SMUs (0.1-degree) resolution. Evening out of extremes happens in two steps from 0.5° grids: (1) when generating 0.5° grid data from the SMU grids and (2) when aggregating from 0.5° grid to NUT2 boundary. This is in contrast to lower averaging of extremes when aggregated directly from 0.1° to NUT2 boundary. The influence of aggregation is more distinct in a region with more simulation cells compared to a region of fewer cells (Table 6.2b) and hence a dependency on geographical region and size of aggregation unit. This is probably because aggregation from high resolution (SMUs) to low resolution such as county or country boundaries reduces the variability in simulated yields. It may also result from heterogeneity of simulated yield patterns and how strongly opposing deviations compensate each other as explained in Porwollik et al. (2017). Yet still, in assessing the effect of aggregation of simulation results of national wheat yield in France, it was demonstrated that aggregation to larger extents affects simulation results and that better results are obtained when predictions are executed at a regional or sub-regional level (Supit and Van der Goot, 1999).

6.4.2.2 Comparison of average WFDEI driven reference yields and average yield forecasts

Figure 6.6a, 6.6b and 6.6c show comparison of WFDEI driven reference yields and North-Gonder NUT2 ensemble average yield forecasts aggregated from 0.5 degree resolution (NUT2_05) for lead-times 0-2 respectively before planting. Yield forecasts show good simulation at lead-0 and lead-1 with (percentage) biases of -72.5Kg ha^{-1} (-4.4%) and -219.1Kg ha^{-1} (-13.4%) respectively. Lead-2 bias is high and unacceptable -841.9Kg ha^{-1} (i.e. -51.6%). This is because climate forecasts (lead-2) do not always span the entire maize growing season when this happens to longer than normal, resulting in crop failures. There exist good agreement in model simulation of mean forecasts and reference yields (i.e. d= 0.6 and d = 0.5 for lead-0 and lead-1 respectively). Fair pattern correlations of r = 0.4 is exhibited in lead-0 and lead-1 forecasts.

Table 6.2: Comparison of observed yields (OBS), NUT2 yields driven by WATCH Forcing data ERA-Interim aggregated from 0.5°×0.5° grid (NUT2_05), aggregated from FAO land use cells of approximately 0.1°×0.1° grid (NUT2) and the respective yield forecasts driven by seasonal climate for lead-times 0 to 2 (i.e. LEAD0 to LEAD2). Results are shown for (a) North-Gonder and (b) Bungoma. Results for Bungoma forecast lead-2 is not shown because forecasts do not span the entire growth. Maize yield units are in Kg ha^{-1}. Asterisk (*) indicate yields that are significantly (at 95% CL) different from the observed. Significance is determined by sigma rule (Equation 2).

(**a**) North-Gondar

	WFDEI			YIELD FORECASTS					
	OBS	NUT2_05	NUT2	05FCST_LEAD0	05FCST_LEAD1	05FCST_LEAD2	FCST_LEAD0	FCST_LEAD1	FCST_LEAD2
MEAN	1581.6	1632.5	2132.7*	1560	1413.5	790.6*	2043.9*	2076.7*	2063.9*
SD	523.3	146.5	155.6	177.2	171.7	77.2	165.4	139.8	123.3
CV (%)	33.1	9	7.3	11.4	12.1	9.8	8.1	6.7	6

(**b**) Bungoma

	WFDEI			YIELD FORECASTS					
	OBS	NUT2_05	NUT2	05FCST_LEAD0	05FCST_LEAD1	05FCST_LEAD2	FCST_LEAD0	FCST_LEAD1	FCST_LEAD2
MEAN	2352	1193.4*	1267.7*	1084.1*	1027.4*	736.2*	1089.6*	1051.0*	
SD	645.2	228.7	289.8	122.4	150.2	114.1	197.1	161.5	
CV(%)	27.4	19.2	22.9	11.3	14.6	15.5	18.1	15.4	

Aggregated from 0.1-degree resolution (NUT2), all the three lead-times show good model simulation and less bias (Figure 6.6d, 6.6e, and 6.6f) for lead-times 0, 1 and 2 respectively. For example the biases range from -56Kg ha^{-1} (-2.6%) to -88.8Kg ha^{-1} (-4.2%). Agreement in model simulation and of forecasts is good in all the three lead-times (i.e. d = 0.6). Good model simulation of inter-annual variability shown

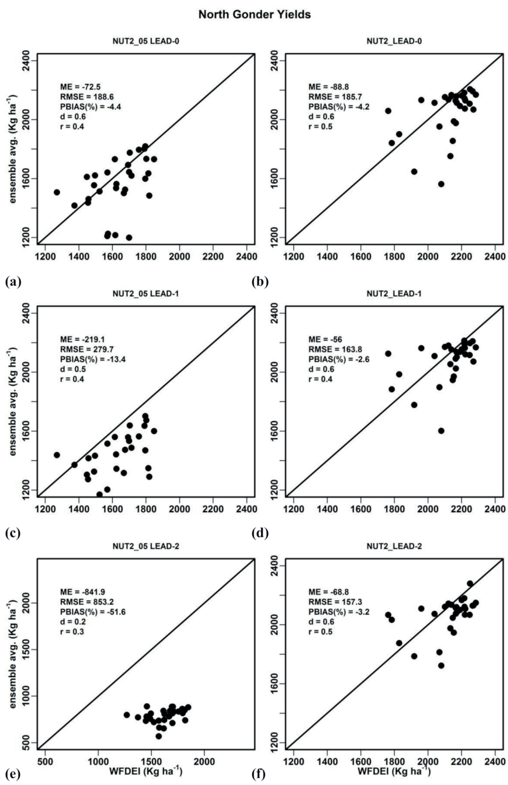


Figure 6.6: Comparison of North Gonder S4 forecast yields, at left aggregated from 0.5° grid resolution (NUT2_05) and at right aggregated from SMU, compared to the WFDEI reference simulation. From top to bottom for lead-0, lead-1 and lead-2, respectively. Each dot is one year (n=30 years). Note the different scales for the vertical axes

by the pattern correlation (i.e. r = 0.4 to 0.5) is exhibited in North-Gonder yields aggregated from 0.1 degree resolution. Similar behavior is found in assessment of North-Gonder biomass (TAGP) simulation (Figure 6.7) even though model simulation of biomass variability is poor as seen from the model agreement index (d) and the pattern correlations (r). The reference WFDEI TAGP simulation is consistently higher (i.e. percentage bias of -2.1, -7.9 and -30.2% for lead-0, lead-1 and lead-2 respectively) than the forecasts. Aggregation from 0.1° resolution has lower biases in all the three lead-times compared to aggregation from 0.5° resolution (see Figure 6.7).

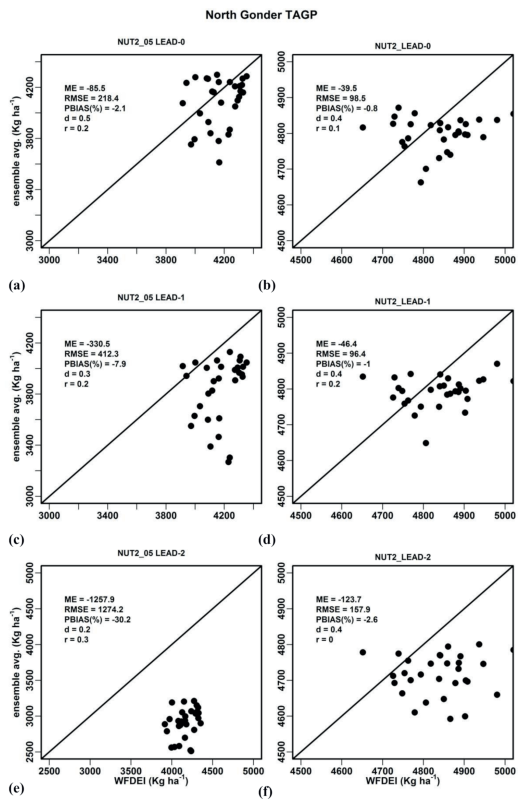


Figure 6.7: Comparison of North Gonder S4 forecast Total above ground biomass (TAGP), at left aggregated from 0.5° grid resolution (NUT2_05) and at right aggregated from SMU, compared to the WFDEI reference simulation. From top to bottom for lead-0, lead-1 and lead-2, respectively. Each dot is one year (n=30 years). Note the different scales for the vertical axes

The fact that (seed) yields are reproduced better than total above ground biomass is contrary to our expectations since flowering and seed filling are subject to a number of additional climate controls (or risks), compared to total biomass; this may be a topic of future research.

Over Bungoma, yield aggregation from 0.5° grid show better model simulation than aggregation from 0.1° grid (Figure 6.8). Good average simulation quantified by ME and RMSE is exhibited for lead time 0 and 1; lead-2 biases (of -38.3%) is high and unacceptable. In all lead-times, the model agreement (d = 0.4 to 0.6) and pattern correlation (r = 0.4 to 0.5) are higher for yield aggregation from the half degree grid. Though aggregated TAGP from half degree resolution (Figure 6.9) show better model simulation compared to aggregation from 0.1° grids, and average biases are small, both show rather poor correlations. In general, better model simulation of

NUT2 North-Gonder yields is observed when yields are aggregated from the higher resolution, native WOFOST simulation unit. The contrast is exhibited over Bungoma where aggregation from the half degree grid exhibit better model simulation. These contrasts may be related to the influence

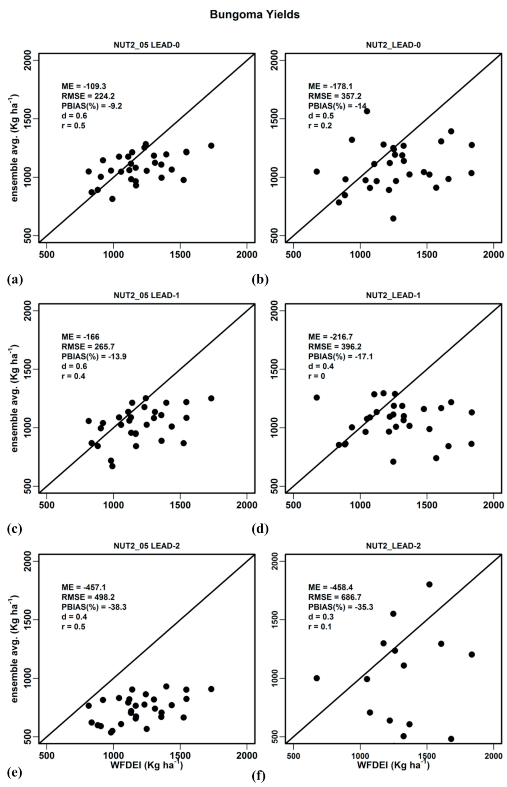


Figure 6.8: Comparison of Bungoma S4 forecast yields, at left aggregated from 0.5° grid resolution (NUT2_05) and at right aggregated from SMU, compared to the WFDEI reference simulation. From top to bottom for lead-0, lead-1 and lead-2, respectively. Each dot is one year (n=30 years). Note the different scales for the vertical axes.

of climate grids. Because Bungoma has only one climate grid, simulation details related to crop-soil-climate continuum is lost but relatively represented over North-Gonder because of its size and the encompassing number of climate grids. Aggregation is therefore influenced by both the size of NUT2 region and the original resolution from which data is aggregated.

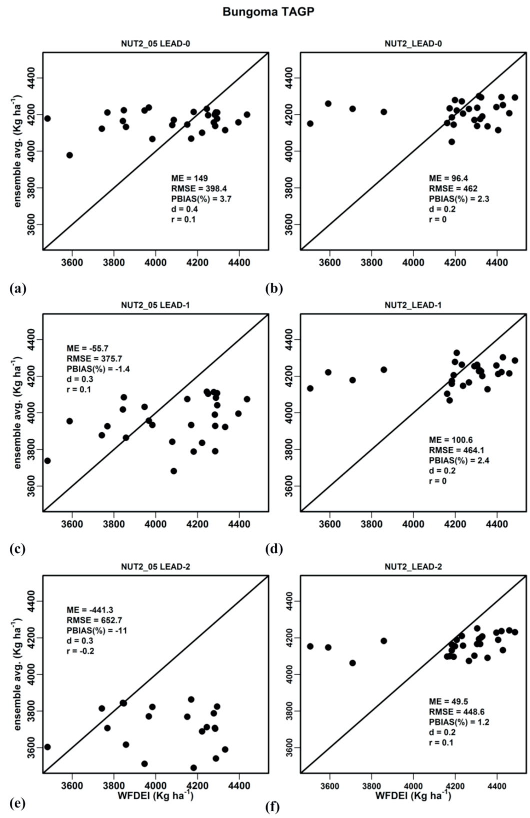


Figure 6.9: Comparison of Bungoma S4 forecast Total above ground biomass (TAGP), at left aggregated from 0.5º grid resolution (NUT2_05) and at right aggregated from SMU, compared to the WFDEI reference simulation. From top to bottom for lead-0, lead-1 and lead-2, respectively. Each dot is one year (n=30 years). Note the different scales for the vertical axes.

6.4.2.3 Influence of spatial aggregation on predictability

Table 6.3 shows comparison of maize yield and biomass predictability aggregated to NUT2 boundary from both 0.5° (Table 6.3a and 6.3b) and 0.1° (Table 6.3b and 6.3d) resolutions. NUT2 yields aggregated from both 0.5-degree and 0.1-degree show comparable below normal (BN) and above normal (AN) predictability in North-Gondar with significant ROCSS of 0.3 to 0.7. Biomass (TAGP) production predictability of BN (ROCSS = 0.2 to 0.5) and AN (ROCSS = 0.3 to 0.6) aggregated from 0.5° resolution is lower than yield prediction. TAGP aggregations from 0.1°resolution show only above normal predictability (Table 6.3b) while below normal production forecast show no predictability. Good predictability in yields compared to biomass (TAGP) may occur because yields are already filtered by only aggregating cells that are harvested while biomass aggregation considers all planted cells.

In Bungoma, yields aggregated from 0.5-degree show higher AN, NN and BN skill compared to aggregation from 0.1-degree resolution whose prediction has no significant skill (Table 6.3c and 6.3d). Above normal yields exhibit higher and significant ROCSS of 0.4-0.5 over three lead times. No matter the resolution from which biomass is aggregated, prediction of TAGP lacks skill.

6.4.2.4 Influence of Aggregation by soil type and crop variety on predictability

Because the study area comprises a diverse climatology and soil types resulting in a number of cultivars, we assessed the predictability of yields aggregated from 0.1 degree resolution to NUT2 boundary based on particular varieties (i.e. yields and biomass for particular varieties are aggregated first before skill assessment) . Table 6.4a and 6.4b show predictability of yields and TAGP for North-Gonder while Table 6.4c and 6.4d show predictability of Bungoma yields and biomass at lead time 0-2. For brevity, we have shown results for only two out of the seven varieties in North-Gonder (i.e. varieties 9 and 3) and Bungoma (i.e. varieties 6 and 3). Both varieties in North-Gonder show good predictability of below normal and above normal yields across all lead times 0 to 2. Below normal forecasts show significant ROCSS = 0.4 to 0.9 while above normal forecasts show significant ROCSS = 0.4 to 0.6 across all forecast lead times. Variety-9 near normal (NN) forecasts only show significant positive skill (i.e. ROCSS = 0.4) at lead-0 while variety-3 forecast has no skill at lead time 0. TAGP is generally less predictable than yields (Table 6.4b). Variety-9 exhibits significant predictability of below normal forecasts (ROCSS = 0.5 and ROCSS = 0.6 for lead times 0 and 1 respectively). Variety-3 TAGP show significant skill for above normal forecasts (ROCSS = 0.4 for lead times 0 only respectively. Below normal and near normal forecasts show no skill.

It is noteworthy, however, that varieties-6 and variety-3 over Bungoma show little yield and TAGP forecast skill (Table 6.4c and 6.4d) when aggregated to NUT2 from 0.1 resolution. While variety-3 show above normal and below normal skill for forecast lead times-0 and -1, but only above normal lead-0 show significant ROCSS = 0.5. Lack of variety-3; at lead-2 predictability is because the variety requires a longer time to mature (see Table-6.1 TSUMS) and as a result does not always grow to maturity with lead-2 climate forecasts. Longer forecasts may perhaps be appropriate. In general, predictability of yields depend on region and not the crop variety. For example, we see good predictability of variety-3 in North-Gonder (Ethiopia) but the same variety has no predictability in Bungoma. Total above ground biomass (TAGP) show less predictability than the yields.

We further assessed the predictability of each variety aggregated to NUT2 region from two dominant soil types i.e. loamy sand-to-sandy loam soils (S3) and sandy loam-to-loamy soils (S4) for North-Gonder and Bungoma (Table 6.5). The ROCSS reported is averaged to NUT2 region for cells with significant skill at 0.1° grids (i.e.

Table 6.3: Predictability of Yield and total above ground biomass (TAGP) at both 0.5° and 0.1° resolution aggregated to NUT2 boundary before predictability skill assessment. Predictability skill scores shown by the ROCSS

(**a**) NORTH GONDAR HALF DEGREE (0.5º) RESOLUTION ROCSS

	YIELD			ABOVE GROUND BIOMASS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.4*	0.3	0.4*	0.2	0.2	0.4*
NN	0	-0.3	-0.1	-0.2	-0.3	-0.2
AN	0.7*	0.7*	0.7*	0.3	0.5*	0.6*

(**b**) NORTH GONDAR 0.1 DEGREE RESOLUTION ROCSS

	YIELD			ABOVE GROUND BIOMASS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.3	0.4*	0.6*	0	0.2	-0.2
NN	0.1	0.4*	0.2	0	-0.1	-0.2
AN	0.6*	0.7*	0.7*	0.4	0.2	0.2

(**c**) BUNGOMA HALF DEGREE (0.5º) RESOLUTION ROCSS

	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.2	0.2	0.5*	0	0.1	0.1
NN	0.3	0.4*	0.1	0.2	0.3	-0.3
AN	0.5*	0.5*	0.4*	-0.1	-0.2	-0.4

(**d**) BUNGOMA 0.1 DEGREE RESOLUTION ROCSS

	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.3	0.1	0.2	0.1	0.4*	0
NN	0.2	0.3	0	0	-0.3	0
AN	0.1	0.1	0.1	0.1	0	-0.1

Table 6.4: Predictability of Yields and TAGP for North-Gonder aggregated to NUT2 boundary by maize variety using ROCSS and it significance. Yields and TAGP are first aggregated to NUT2 before skill assessment.

NORTH GONDAR (0.1° RESOLUTION)

(a) AGGREGATED YIELDS TO NUT2 REGION BY VARIETY

	VARIETY-9			VARIETY-3		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.7*	0.7*	0.4*	0.5*	0.9*	0.6*
NN	0.4*	0.3	0.2	0	0.1	0.1
AN	0.5*	0.4*	0.5*	0.4*	0.4*	0.6*

(b) AGGREGATED TAGP TO NUT2 REGION BY VARIETY

	VARIETY-9			VARIETY-3		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.5*	0.6*	0.3	-0.2	-0.1	-0.2
NN	0	0.4*	0	-0.2	0.2	0
AN	0.1	0.2	0.5*	0.4*	0.3	0.1

BUNGOMA (0.1° RESOLUTION)

(c) AGGREGATED YIELDS TO NUT2 REGION BY VARIETY

	VARIETY-6			VARIETY-3		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0	0	0.3	0.2	0.2	0
NN	0.3	-0.3	0	0.2	0	0
AN	0.1	0.2	0.4*	0.5*	0.3	0

* shows significant skill at 95% level

(d) AGGREGATED TAGP TO NUT2 REGION BY VARIETY

	VARIETY-6			VARIETY-3		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN						
NN						
AN						

SMU skill assessed before aggregation to illustrate predictability at high resolution). To be useful to farmers, we also report the percentage of 0.1-degree cells within the NUT2 region that show significant skill (shown in brackets in Tables 6.5 and 6.6). We consider varieties-9 and -3 from North-Gonder. Variety-9 is more predictable from S3 than from S4 soils with below normal forecasts showing lead-0 and lead-1 significant ROCSS of 0.5, in addition 30% of the SMUs show significant skill scores. Variety-9 above normal forecast show good predictability in all lead times (ROCSS = 0.4 to 0.5) and 30% grids show significant scores. Aggregated over S4 soils, all categories of forecasts show insignificant ROCSS (Table 6.5a). Variety-3 has better predictability in both the S3 and S4 soils; it also shows higher percentages of SMUs having significant ROCSS (i.e. ROCSS between 0.5 and 0.7, and percentages of 48% to 100%) in S3 soils. S4 soils show significant prediction skill score in 22% to 100% of grid cells.

We consider only the dominant S3 soil type which covers over 93% of planted cells in Bungoma (Table-6.6). Prediction of other varieties with respect to dominant soil types is shown in table D.12. Aggregations by soils show no predictability in Bungoma. Predictability differs with soil types irrespective of maize varieties. This suggests dependence of yield predictability on soil-cultivar-climate interactions and the need to assess forecast skill at higher resolutions in which high spatial variability in both soils and climate are captured.

6.4.2.5 Prediction skill upscaling from high resolution grid (i.e. 0.1° grid)

We report above-normal (AN) and below-normal (BN) prediction skill averaged over cells with significant ROCSS plus their percentages. We show the skill of yield or biomass aggregated upwards to NUT2 boundary considering all varieties in each region. In Table 6.6, we report the predictability of dominant crop varieties in each region for illustration purposes (i.e. varieties-9 and -3 for North-Gonder; and varieties-6 and 3 for Bungoma). In general, North-Gonder exhibits higher predictability than in Bungoma even for the same crop variety (e.g. variety-3). At this high resolution there exists almost no predictability in Bungoma. This finding agrees to that in the section-3, i.e. prediction at the 0.1-degree grid is higher in the larger NUT2 area than in the smaller one, perhaps related to the lack of detailed, small scale interactions between crop, climate, and soils.

In North-Gonder, average of skill in grids that show significance for individual varieties result in higher skill than with all the varieties (NUT2) considered (see Table 6.6a, and 6.6b). ROCSS is greater than 0.4 except for lead-2 crop production forecasts and variety-3 TAGP over North-Gonder. Considering individual varieties, lead-times that possess significant skill are greater than 42%. Even though ROCSS are still comparable for NUT2 region, the percentages of cells with significant skill are lower. Averaging over NUT2 boundary acts to degrade prediction information that would otherwise be useful for near farm-scale prediction.

Table 6.5: North Gondar NUT2 Yield and TAGP predictability aggregated by maize variety and soil types at FAO soil mapping units (0.1°resolution).Predictability is shown using ROCSS averaged over grid cells with significant skill scores. In brackets are percentage of grid cells with significant ROCSS. Soil type S3 means “LOAMY SAND-to-SANDY LOAM”, and S4 means “SANDY LOAM-to-LOAM soils

NORTH GONDAR (0.1° RESOLUTION)

(a) AGGREGATED YIELDS TO NUT2 REGION BY SOIL TYPE

	VARIETY-9							VARIETY-3					
	S3 SOILS			S4 SOILS				S3 SOILS			S4 SOILS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2		LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.5* (30%)	0.5* (30%)	0.1	0	0.1	0.1	BN	0.7* (48%)	0.6* (100%)	0.1	0.7* (22%)	0.5* (100%)	0.2
NN	0.1	0.2	0	0	-0.2	0	NN	0	0	0.1	0	0	0.2
AN	0.4* (30%)	0.4* (30%)	0.5* (30%)	0.3	0.2	0.2	AN	0.5* (48%)	0.6* (48%)	0.5* (96%)	0.5* (22%)	0.6* (22%)	0.4* (78%)

(b) AGGREGATED TAGP TO NUT2 REGION BY SOIL TYPE

	VARIETY-9							VARIETY-3					
	S3 SOILS			S4 SOILS				S3 SOILS			S3 SOILS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2		LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.65* (30%)	0.5* (100%)	0.2	0.3	0.5* (100%)	0.2	BN	0.1	0.4*(45%)	-0.1	0.1	0.4*(22%)	-0.1
NN	0.16	0.45*(70%)	0	-0.2	0.3	-0.1	NN	-0.1	-0.1	0	-0.1	0	-0.1
AN	0.37(30%)	0.47*(30%)	0.51*(30%)	-0.2	0	0.1	AN	0	0.4*(48%)	0.1	0	0.4*(22%)	-0.1

* shows significant skill at 95% level

Table 6.6: Predictability of NUT2 TAGP and maize yield per cultivar variety at FAO land use cells (i.e. 0.1° resolution). Predictability is shown using ROCSS averaged over grid cells with significant skill scores. Skill ia assessed before aggregation of ROCSS to NUT2. In brackets are percentages of grid cells with significant ROCSS. First 3 columns are results over all varieties in a NUTS2 region, others for a specific variety in that region.

(a)	ANALYSIS AT SMU RESOLUTION NORTH GONDAR YIELDS								
	NUT2			VARIETY-9			VARIETY-3		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.6(15%)	0.6(29.7%)	0.1	0.5(42.1%)	0.5 (42.1%)	0.2	0.6 (44%)	0.5 (100%)	0.1
NN	0	0.3(3.3%)	0.1	0.2	0.2	0.1	0	0	0.1
AN	0.4(27%)	0.5(27%)	0.4(22%)	0.4 (42%)	0.4 (42.1%)	0.5 (42%)	0.5 (43.6%)	0.6 (44%)	0.5 (56%)
(b)	NORTH GONDAR TAGP								
	NUT2			VARIETY-9			VARIETY-3		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.5(7%)	0.5(22%)	0.5(2.3%)	0.6(42.1%)	0.6 (100%)	0.1	0.1	0.5 (40%)	-0.1
NN	0.5(2.8%)	0.5(7%)	0.4(5%)	0	0.5 (73%)	-0.1	-0.1	0	-0.1
AN	0.5(6%)	0.4(17%)	0.4(6%)	0.5 (42%)	0.5(42%)	0.4(42%)	0	0.4 (44%)	0.1
(c)	BUNGOMA YIELDS								
	NUT2			VARIETY-6			VARIETY-3		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0	0.1	-0.2	0	0.1	-0.3	0.2	0.2	-0.1
NN	0.5(2.2%)	-0.3	0	0.5(2.2%)	-0.3	0	0.1	0	0
AN	0.1	0.2	0.2	0.1	0.2	0.2	0.5(50%)	0.3	0.1
(d)	BUNGOMA TAGP								
	NUT2			VARIETY-6			VARIETY-3		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0.3	0.1	0	0.3	0.4(2.2%)	0	0	0.2	-0.1
NN	0.1	0	0	0.1	0	0	0.1	0.3	0
AN	0	-0.1	0.2	0	0	0.2	0.2	0.3	0.1

6.5 Discussion

6.5.1 Setup

Crop models are basically meant for site based studies; for a regional study, detailed actual soil characteristics, detailed actual crop (let alone variety) characteristics and climate observations at field resolution are not available. The method we used in this study was the best possibility practically with soil and climate data. The setup may change in future similar studies especially with regard to climate data. For example, climate predictions from the European Centre of Medium range weather Forecasts (ECMWF) used in this study have a horizontal grid resolution of $0.75^{\circ}\times0.75^{\circ}$ but future forecasts will be availed at $\approx$30 km grid resolution. There also exist high resolution gridded climate observations in the East African countries of Tanzania, Kenya and Ethiopia, such as the Centennial trends precipitation dataset for the Greater Horn of Africa (CENTRENDS) for the period 1900-2014 at 0.1° horizontal resolution, produced by the University of California at Santa Barbara, Climate Hazards Group and Florida State University (Funk et al., 2015); the Climate Hazards Group InfraRed Precipitation and Stations (CHIRPS) at 5km resolution starting from 1981 to near real-time (Funk et al., 2015) is now operationally used in the region . With efforts to grid other variables, the observations would replace the pseudo observations (WFDEI reanalysis) used in this study and hopefully improve simulations of reference yields. Even so WFDEI has been validated globally and used in many studies over eastern Africa.

6.5.2 Comparison of observed and simulated yields

The differences between OBS and simulated mean yields and their standard deviations occur because some yield-limiting factors inherent in OBS are omitted in our model simulations as they are not incorporated in the WOFOST model. This includes factors such as effects of weeds and pests. Furthermore, water conserving techniques that are generally applied in the dryer regions are not considered. This may lead to an over estimation of dry spells stress in these regions. The fact that a simple tipping bucket approach is used in our WOFOST version instead of a more detailed soil moisture module may contribute to this overestimation. The simple bucket soil moisture balance approach amplifies sensitivity to drought conditions in WOFOST. Fertilizer application is kept constant in our simulations but in reality, it may vary from one year or season to another hence influencing the observed yields. Difference may further be compounded with the fact that only one season is simulated in a single year yet some regions have double cropping seasons.

6.5.3 Influence of spatial aggregation on predictability

Many large scale crop models simulate yields at resolutions of approximately 0.5° almost equivalent to climate data resolutions used in this study. WOFOST setup simulates yields and biomass production at $0.1^{\circ}\times0.1^{\circ}$ a resolution that may be useful to farmers climate service needs. We assessed maize yield predictability at both half degree resolution and national boundaries for Eastern Africa in Ogutu et al. (2018). The assessment is important in forecasting large scale crop performance covering regional, national and sub-national boundaries. Furthermore, official agricultural yield statistics (observed); crop production forecasts or outlooks are normally issued for national or sub-national (NUT2) regions. Irrespective of the spatial resolutions of crop model simulation units, crop production therefore needs to be aggregated to enable provision of outlooks or a comparison to the observed. In this study, we assessed the influence of aggregation from 0.1° to 0.5° resolutions and on to NUTS-2 level using two case-studies over two regions: one in the Northern part of the study area (North-Gonder) and another in the equatorial regions of East Africa (Bungoma), using a criterion established in section 6.3. These are regions of good climate predictability and maize yield predictability. They also exhibit different climate characteristics, topography, and cropping seasons, but also have good time series of observed data. For farmers to benefit from this research, we also assessed predictability at WOFOST simulation unit in this study i.e. FAO soil mapping units (a resolution of approximately 10 km ×10 km grid). We assessed both simulation skill and predictability of yields and biomass, predictability from different soil types and different crop varieties.

Poorer predictability in Bungoma compared to North-Gonder may result from the fact that within its one single climate grid cell a large range of altitudes occur (Figure6.2) , while many of the 0.5° grid cells in North-Gonder are rather homogeneous in altitude i.e. Bungoma spans almost 2000m altitude in its single 0.5°cell, in N-Gonder only 2 cells out of 28 show the same range, all others much less and more than 10 cells show no altitude variation at all. Therefore, fine resolution climate characteristics that would affect variability in simulation and predictability at the higher 0.1° resolution are not captured in the crop model WOFOST, especially rainfall that has higher spatial and temporal variability. Lack of spatial variability in climate variables over Bungoma are evidenced by the coefficient of variations of rainfall rate, global radiation, minimum temperature and maximum in the supplementary Figures C.15 to C.18 respectively. In all instances, CV of each ensemble member is nearly equal to that of reference WFDEI except for maximum temperature CV that is higher but still no more than 2% above the reference.

It means therefore that higher resolution forcing climate data may improve the predictability by capturing more realistic climate characteristics and spatio-temporal variability. In contrast, North-Gonder is a larger area, covered by a number of climate grids resulting in more spatial details being captured over several grids hence better simulation. Aggregation over a larger area such as North Gonder may improve predictability over a regional extent. Similar setup of crop simulation studies, i.e. at modeling units then aggregating to a larger area has been operationally employed to estimate yields over Europe using the CGMS of the European Union Joint Research Center (EU-JRC) (Boogaard et al., 2013; Supit and Van der Goot, 1999). Even though good simulation has been attained when input data is aggregated over larger area, for example in simulation of crop phenology over Germany using AFR-

CWHEAT2 decreasing the spatial extent of input data aggregation further improved simulations (Van Bussel et al., 2011). This demonstrates the value of detailed input data in crop simulation. The method employed in this study is reasonable especially in a region of diverse climates (and hence crop calendars), soil characteristics, and surface topography like eastern Africa.

There is no consistent influence of forecast lead-time on yields or biomass predictability aggregated to NUTS-2 boundary. Perhaps because our main focus is predictability of anomalies rather than absolute values and the statistics employed may not be practical to assess the simulation of absolute yields. Biases in simulated yields and biomass change with forecast lead time, but are still either above or below normal when aggregated over a larger area.

6.5.4 Aggregation by crop variety and different soil types

Yield responses to climate have been found to vary over different soils due to varying rooting depths, plant available water capacity, hydrologic conductivity, among other properties that affect yields . But in this study, we do not get significant variations in yields over different soil types, irrespective of crop varieties. This is probably because the dominant soil in the study regions have almost similar available water capacities. The maximum soil depths are also fixed for all soil types in addition to the simple soil water balance in WOFOST. These would not allow enough variability in yields over different soil types. We recommend that including detailed soil water balance processes may result in variability in yields over the different soils. This could consequently result in better predictability over different soils.

For each variety, simulation and predictability of yields are better than for biomass. This may be because of the high sensitivity of biomass production to sunshine, solar radiation. The inter-annual variability in solar radiation during the crop growing season in the present tropical study region may not be significant compared to variability in rainfall (see supplementary information in Figure C.15 and Figure C.16 for rainfall and radiation respectively). In WOFOST, the Harvest Index (HI) is not constant from one year to another but is rather influenced by weather during certain stages of the crop growing season. For example, rainfall during grain filling stage results in higher HI and hence higher yields but will not have a similar influence on biomass. Rainfall for example is already known to have higher spatio-temporal variability than other weather variables and would significantly influence HI. The harvest index therefore varies annually and differently among the crop forecast ensembles leading to high variability in yields. This may have contributed to better predictability in yields than biomass. This is clearly seen in North-Gonder region of Ethiopia as opposed to Bungoma where there exist less variability in growing season weather (shown in Figure C.15 to C.18 in the appendix) resulting in less variability in HI and consequently less yield variability.

6.6 Conclusions

This study examines the influence of aggregation on seasonal climate driven maize production forecast. A field scale crop model (WOFOST) is configured to simulate maize production over East Africa at a high spatial resolution of 0.1° grid (SMU) corresponding to input soil data grid. Climate grids are however coarser; implying that all SMU within a climate grid would have similar climate characteristics. This resulted in crop simulation missing the detailed small scale crop-soil-climate interactions; especially so in a region with steep rainfall variability such as East Africa. Crop production forecasts are generally issued over national (country) or sub-national (NUT2) regions that are larger than most crop model simulation units. Simulations and predictability at such regional extents are relevant for policy interventions. In this paper we have assessed the influence of crop model output aggregation over NUTS-2 boundaries by soil types, and by crop varieties over two case study areas (i.e. North-Gonder in northern Ethiopia, and Bungoma in the equatorial region of Kenya). We also assessed predictability at the high resolution (0.1° grid) relevant to farmer decisions with respect to soil types and crop varieties. We conclude that:

Influence of aggregation on average yields is region dependent. In this study good yield simulations from NUTS-2 areas of larger extent (such as North-Gonder) are close to observed values, compared to larger biases from smaller spatial extents (such as Bungoma). Mean yields do not vary much with forecast lead-times. Yields aggregated to NUTS-2 regions from 0.5° resolution show good simulation compared to that aggregated from the fine 0.1° simulation units. Lack of good simulation in for example Bungoma may be related to climate characteristics, i.e. in smaller areas, detailed climate characteristics are not captured because of coarse resolution of the climate inputs. The unit of aggregation influences simulation skill at sub-national and presumably national boundaries.

Poor simulation of biomass production compared to yields may be related to the length of growing seasons because yield aggregation is already filtered by harvested cells (at maturity) while TAGP harvests are not filtered i.e. there will be TAGP as long as a plant grows.

Yields aggregated from 0.5° grid exhibit higher yield predictability of above-normal and below-normal forecast than yields aggregated from 0.1-degree grid.

Predictability depends on soil types; this suggests dependence of yield predictability on soil-cultivar-climate interactions and the need to assess forecast skill at higher resolutions in which high spatial variability in both soils and climate are captured. Further research should use a more detailed climate-soil-moisture data.

Yield and biomass prediction and/or simulation over NUTS-2 region are variety and region dependent. Prediction of a suitable variety may be good in one region but not another, for example the predictability of Variety-3.

Acknowledgement This study was financially supported by the EUPORIAS project (European Provision of Regional Impact Assessment on Seasonal-to-decadal timescale); grant agreement No. 308291 funded by the European Commission in the Seventh Framework Programme for Research. We acknowledge the assistance offered by Wietse Franssen and Ham van der Henk regarding coding and Oracle SQL respectively.

Chapter 7

General Discussion

7.1 Research questions

This thesis explored the use of dynamic seasonal climate forecasts and a crop model to predict water-limited yields in the Greater Horn of African countries of Ethiopia, Kenya and Tanzania. A region characterized by vulnerability to the year-to-year variability in climate shocks. The basis of this research is that skillful seasonal climate forecasts should translate also into skillful yield forecast in the climate-impact modelling chain. The current practice to issue crop performance outlook in the GHA is by expert opinions depending on climate forecasts issued over broad areas of the study region. The climate forecasts have been based on established statistical relations between local climates and global teleconnections (such as SSTs among others) in the climate system. In the recent past however, developments in dynamically generated Global Climate Model (GCM) forecast have resulted into improvement in skill surpassing the skill in statistical forecasts in some regions of the globe. This provides a potential for use in providing objective forecast or outlook of crop performance. The technological requirements, complexity, and huge data requirement in using climate forecasts and crop models may hinder application of such a system in some regions. Therefore a simpler relationship between climate characteristics during various stages of crop growth and final yields is assessed with the aim of providing a means of using less complex statistical methods to forecast yields based on weather evolution during crop growth stages. Under usual circumstances, official yield statistics in the GHA countries are reported at nation administrative boundaries or sub-national level and even though yield simulation may be at high resolution, there was a need to assess the impact of aggregation by spatial boundaries, by soils and by cultivars. The results of this study are arranged in chapters with chapter 1 giving a background to the study, chapter 2 describes crop model setup for yield simulation, chapter 3 reports on skill of seasonal climate forecast in the region, chapter 5 reports of crop model simulation and yield forecasts, chapter 4 reports on the relationship of growing season climate characteristics and predictability of significant indicators, while chapter 6 assesses the influence of simulated yield aggregation on forecast skill.

Four research objectives were formulated to guide this study and research questions answered in chapters 3 to chapter 6. However in the following paragraphs, we discuss and summarize findings corresponding to each research objective.

1. *Assess the skill of GCM and bias corrected seasonal climate forecasts over East Africa via comparative analysis of hindcasts with observational data for the period 1980 to 2011*

   This objective addressed issues related to ECMWF system-4 (*S4*) seasonal climate forecasts by answering four research questions i.e.

   - *How well does the S4 simulate climatology and seasonality of the variables relevant to agricultural impacts modelling over East Africa?*
   - *How does climate forecast skill for these variables at various lead-times compare for each season relevant to the sub-regions?*
   - *Does bias correction either improve or adversely affect skill for lead-times, seasons and geographical forecast units assessed?*
   - *How well does S4 simulate anomalous wet and dry years associated with positive or negative ENSO phases?*

   We answered the questions by comparing climate reforecasts to gridded observations in seasons that are relevant to cropping seasons in eastern Africa i.e. March–May, June–August, October–December (MAM, JJA, OND). We focused on rainfall *tp*, surface air temperature *tas* and downwelling short wave radiation *rsds*. The primary reference data is the WFDEI which is based on ERA–Interim reanalysis, but the reanalysis share the same atmospheric model (but different versions) as the *S4* forecast system. This means that there exists an inter-relationship between the two. To assert the robustness of our results, we used other independent "observed" datasets i.e. the 0.1° African Rainfall Climatology (Novella and Thiaw, 2013) version–2 (ARC2), the University of Delaware (UD11) near–surface air temperature (Willmott and Matsuura, 2001), and radiation data from NASA/GEWEX Surface Radiation Budget (Zhang et al., 2013) release–3 (SRB3). To address the first research question, we used bias (i.e. difference between mean ensemble forecast and the observed) to asses how well *S4* simulates climatology and pattern correlation to assess seasonality and interannual variability. The second research question is addressed by use of the Relative Operating Curve Skill Score (ROCSS) and the Ranked Probability Skill Score (RPSS) to asses forecast skill as a function of forecast lead time, season and geographic sub-region of the study area. The same verification metrics are used for both bias-corrected and raw model simulations. The ROCSS is used to assess simulation of anomalous years. We answer these four questions starting with rainfall, temperature and shortwave radiation in the following paragraphs.

S4 simulated well the interannual variability, spatial patterns and structure of rainfall in all seasons but bias characteristics depend on season, and the validating data set. Change in rainfall biases with increasing forecast lead-time in MAM are unsystematic for example, forecasts initiated in November or December of the previous year (lead months -4 and -3 respectively) show less biases. We argue here that the status may have resulted from model drift or due to the stability of ENSO circulation in November–December but noting that no ENSO influence in MAM rainfall has been evidenced (Camberlin and Philippon, 2002). Same unsystematic change in rainfall biases is also seen in Ethiopia in JJA season.

Cold biases with a persistent structure dominate temperature forecasts i.e. colder temperatures in *S4* forecasts than the observed. This feature is irrespective of the verification data use i.e. found whether validated against WFDEI or UD11. The structure of biases persist over longer forecast lead-times that may be because local influences such as land-use, changes in soil moisture and orography are not well represented in the model. Since *S4* has a course grid resolution it may be that the local features are not well represented. We hypothesize that biases may be lowered by higher resolution climate forecasts or by a prognostic downscaling method that takes care of the relationship between large scale features and local climate.This may perhaps be improved in future similar research since ECMWF forecasts have since increased the resolution of their operational forecast system (SEAS5) (Johnson et al., 2019). Also, there is also a successor to WFDEI currently available from C3S, the WFDE5 (Cucchi et al., 2020) produced by aggregation of ERA5 as opposed to interpolation of ERA-Interim reanalysis to WFDEI grid. It has higher spatial variability and higher variable correlations to FLUXNET (Baldocchi et al., 2001) observations. Their performance in crop models would need to be assessed. Temperature biases seem to follow elevation i.e. warm biases in higher grounds with elevation $\geq$ 1500m and cold biases in lower elevation areas. Knowledge of region specific biases in model simulation are important for it then dictates how they can be applied in impact studies. Temperature forecast show significant anomaly correlations with the observed irrespective of the verifying data set indicating good simulation of seasonality or interannual variability. Where biases are high (up to 3° C) in some areas, post processing techniques such as bias correction are needed to improve their suitability for use in impacts modelling. The best methods of reducing the biases would have to be investigated.

Conspicuous differences in the magnitude and nature of biases in downwelling shortwave radiation that is dependent on both verifying data and topography prevail especially in the June–August season in the Ethiopia highlands. There exist high positive, and low negative bias against WFDEI and SRB3 respectively. WFDEI is produced from the ERA-Interim reanalysis. In production of the ERA–Interim, atmospheric part of the model IFS assimilates observed data from the World Meteorological Organizations' (WMOs) ground, sea and upper–air networks (Dee et al., 2011) and produces a series of constrained daily best possible historical state of the climate at IFS resolution. Even though the model is kept as close as possible to assimilated observation, drift and chaotic processes produce biases. Also, the observational network of ground and especially upper air network is sparse in the GHA region, also reducing potential accuracy of the assimilation. Precipitation and temperatures simulations are bias corrected against *in situ* observations (CRU–TS3.1 (Harris et al., 2014) in the version of WFDEI used) to produce the WFDEI but the radiation data is not corrected against elevation (Weedon et al., 2014). Topographic effects that include terrain orientation, slope, solar illumination angle are known to determine downwelling shortwave radiation form space (Zhang et al., 2015)and could also contribute to the nature of biases in both the low elevation regions and mountainous areas. This may be accentuated by cloudy conditions in Ethiopian highlands in JJA. Atmospheric environment that include for example the aerosol load and their extinction properties may also play a role. However, diagnosing causes of *rsds* is not part of this study and may be an area for further research. Significant anomaly correlations exist in lead-0 against all reference data sets but reduce with longer lead-times similar to precipitation skill, and sharing perhaps some causal relationships in common parameterizations.

In assessing predictability, the RPSS is found to be insufficient as a measure of predictability due to sensitivity to less predictable middle-tercile forecasts (near normal). We use ROCSS to evaluate forecast skill of either upper or lower terciles with the result that *tas* is most predictable followed by *tp* and *rsds* in that order. The pattern of grid cells having significant ROCSS are nearly similar across all verification data sets. Near–normal forecasts are less predictable agreeing with other studies in Kharin and Zwiers (2003) and Van Den Dool and Toth (1991). Lead–times of useful forecast skill however vary with season and geographical region reiterating the need to verify model simulation skill at a local level. It is however not known what a suitable spatial scale is best to verify forecasts, in general though, a single skillful cell surrounded by no-skill cells has little meaning. Anomalous wet/dry years in the GHA are largely driven by El-Niño and La Niña conditions, in interaction with amongst others the IOD (Goddard et al., 2001; Saji and Yamagata, 2003; Black, 2005; Owiti and Ogallo, 2007; Owiti et al., 2008), and local features such as inland lakes (Song et al., 2004; Thiery et al., 2015) also exert influence. As such, some historical anomalous years [i.e. 1982/1983, 1997/1998 (wet years); 1984/1985,1999/2000 (dry years)] are captured by the forecast system and not others. However, the GCM forecasts offer an advantage over the traditional statistical forecast in some regions and season. For example, JJA forecasts in northern Ethiopia show skill beyond two lead-times where as Nicholson (2014); Korecha and Barnston (2007) found no more than 2-months lead time in prediction of JJA rainfall in the GHA.

GCM forecasts have also be found to perform better than statistical forecasts in other regions of the world (McIntosh et al., 2007).

2. *Assess the important growing season climate characteristics or indices that influence maize yield predictions, and their predictability*

This objective was based on the hypothesis that rainfall and temperature are the main drivers of rainfed crop yields, but in as much as yields are determined by the amount of rainfall and average temperature during a crops' growing season, the behavior and evolution of these climate variables over the growing season may play an even bigger role especially when favorable or extreme weather events occur during specific critical stages. This was assessed in two geographical regions in the study area where both skillful yield and climate forecast have been obtained in chapter 3 and **??**. The ambition was to:

- *identify rainfall and temperature characteristics (indicators) that explain yield variability (from WOFOST WFDEI driven yields)*
- *Asses their predictability, and uncertainty in prediction as a function of maize phenological stage and climate forecast lead-time*

We used rainfall indicator that describes daily rainfall rate (R_R); even rainfall distribution-evenness parameter (E_R); uneven random rainfall events-unevenness parameter (E_R), and time of occurrence (Ad) i.e. whether early, middle or late in the season. We used temperatures as indicators i.e. daily maximum temperature (TX), daily minimum temperature (TN), average daily temperature (TG), and derived indicators (GDD and KDD). Two regions of the study area in northern Ethiopia (North-Gonder) and equatorial Kenya (Bungoma) based on good climate and yield prediction skill from chapter 3 and chapter 5 skill are selected for this case study. While we use the mean of climate variables (and therefore indicators) over each region, yields are aggregated using Equation 4.1. The relationship between yield and climate indicators are explored considering the entire growing season, vegatative stage, ±10-days around anthesis and in the reproductive stage.A range of forecast verification metrics sampling different attributes were used. The general finding in this is that depending on geographical region or phenological stage under consideration, climate indicator that explain yield variability differ and it cannot be generalized for the two region. Notable however is that some rainfall and temperature indicators (not necessarily the same in each phenological stages) explain yields in North-Gonder (Northern Ethiopia). During anthesis and reproductive stages, temperature indicators explain yield variability in both case study regions. Yield variability is not explained by rainfall in Bungoma probably because of sufficient soil moisture or rainfall. Water availability during anthesis and in the reproductive stage are found highly explanatory of grain yields (Armstrong, 1999; Butts-Wilmsmeyer et al., 2019). Since both water and temperature during anthesis and reproductive stage are known to affect yields, we conclude that sufficient water requirements could be the only reason why temperature indicators explain yields but not rainfall indicators. This agrees to findings in other studies that maize yields are sensitive to climate variables but with highly geographical region dependent sensitivities (Ray et al., 2015; Huang et al., 2015, 2017; Zhao et al., 2015; Karunaratne and Wheeler, 2015; Kamali et al., 2018; Iizumi et al., 2013). Identification of variables that explain yields in each region and growth stage makes it possible to forecast yields with the same by statistical methods, mitigating the need to use complex climate forecast and crop models. Their predictability (or lack of) allows for probable yield predictions depending on phenological stages. Predictability in this study is assessed at lead times before planting. But for application purposes, it is possible to obtain forecasts of the indicators at shorter lead-times. Mitigation measures against poor yields may then be developed for shorter periods making them less expensive to farmers.

Thus predictability of the indicators (those that explain yields) during growing season, phenological stages, and forecast lead-months were assessed for the two case study regions. Generally, growing season rainfall distribution indicators (R_R, E_R) are more predictable than U_R, and TN, TX, TG are predictable than KDD. GDD is however not predictable in both Bungoma and northern Ethiopia. Noteworthy is that not all variables are predictable with same skill nor forecast terciles nor forecast lead times, i.e. predictability varies with region, tercile, crop growth stage and forecast lead-time. For example, R_R, E_R, TN, TG, KDD above-normal forecast are predictable in northern Ethiopia three lead-months before planting but only lead-2 forecast of TN is predictable, also below-normal forecast of Ad is predictable but not AN forecasts. While in Bungoma, only temperature indicators are predictable at lead-0 and lead-1. This is despite skillful rainfall forecasts in (Ogutu et al., 2017) from which the indicators are derived illustrating the importance of rainfall indicators rather than the average over a season. Variability of skillful/non-skillful predictions of indicators varies with the crop phenological stages, regions and forecast lead-months and should be assessed as such. As does the uncertainty in prediction of these indicators. Assessing a single attribute with more than one measure however has challenges, for example we use spread-to-error (SprErr) ratio and the rank histogram to assess uncertainty but in some instances,they do not lead to similar conclusions. For example, North-Gonder GDD forecasts show certainty in some stages when SprErr is used ($0.5 \leq \mathrm{SprErr} \leq 1.5$ in our case) but consistently shows non-uniformity of the rank histogram (uncertainty) in some stages and lead-times despite GDD terciles being predictable in northern Ethiopia. This observation is not limited to GDD but also other variables thus questioning the value of using many measures to evaluate a single forecast attribute. Predictability of some of the indicators with good forecast lead-times however offers the potential of early prediction of yields by statistical methods thus eliminating the need of complex climate-crop model yield simulation.

Even though with simulation biases, we found potentially useful skill in predictions of rainfall, temperature and shortwave radiation with at least 2-months lead time before start of cropping seasons in different regions of the study area. The next was to apply these forecast into a crop model and evaluate how well yields are simulated. Thus the next objective. The new ECMWF operational seasonal forecast system (SEAS5) with higher number of ensemble members in SEAS5 can give better performance statistics but the effect of ensemble members need to be ascertained. This is because in assessing the effect of ensemble size on probabilistic verification, Kumar et al. (2001) found that 10-20-ensemble members is sufficient to ensure average skill close to the expected from infinite ensemble size. But the generally higher skill of Multi Model Ensembles may in future become advantageous since the forecasts have become available in harmonized fashion on C3S (C3S, 2017). Their use in crop yield forecasting could be explored in future research.

3. *Assess maize yield predictive skill of a crop simulation model through both baseline and hindcasts validation for the period 1980 to 2011*

   This objective was addressed in chapter **??**. We first evaluated WOFOST simulated historical maize sowing dates and yields then we assessed skill of probabilistic yield forecasts based on ensemble GCM climate forecasts. The aim of which was to identify lead times and areas of potential pre-season yield forecasting. We assess how well hindcast of maize yield emulates yield anomalies due to interannual climate variability.

   The first step in this part of the thesis was to set up WOFOST (chapter 2) to simulate yields over a regional scale given that WOFOST is a model meant for a field scale area. This required a databases with quantitative information on crop, soil, and climate characteristics. Quantitative because over reliance in expert knowledge in agro-ecological engineering may result in biased designs (De Ridder et al., 2000) i.e. ideas that do not correspond to the experts' perception may be dismissed. First, the crop varieties identified in this study may differ from the actual varieties planted in the study region, but the yield simulations are found to compare well with the observed. Crop calendars have been found to be in agreement with other large scale calendars such as the FEWSNET calendar (Figure C.10). Though the planting dates are constant, optimized on maximum average yields, the dates fall within those in FEWSNET.

   This is true for both the simulated sowing and harvest dates. It is noted however that actual farmer planting dates may differ depending on other factors that are not climate related. This can be a subject of further research and may be accomplished by working with farmers and using the actual planting dates then compare with simulated results. Deviations from planting dates in this study would however not differ extremely because planting in rain fed agriculture is driven by rainfall onset to a larger extent (apart from non-climatic factors) at the start of a season; and our dates are climatology based and fixed.

Harvest dates are determined by interactions between crop cultivars, soils and climate (and depends on TSUMS). Differences between reference harvest dates and predicted dates do not exceed 10-days illustrating acceptable predictions (Figure 5.2).

We found good anomaly correlations (r=04 to 0.6) between simulated yield anomalies and WFDEI rainfall anomalies indicating that climate variability drives maize production. There is almost nil correlations (i.e. r=0.02-0.2) between observed yields (both FAO and National statistics) and WFDEI rainfall anomalies (anomalies are shown in 5.4) indicating climate or rather rainfall is not the only factor influencing yields (actual yields). Validating observed historical yields and WOFOST simulation, mean differences of between 600-700Kg ha^{-1} were found for Kenya and Ethiopia but lower (i.e. $\approx$ 200Kg ha^{-1}) for Tanzania between the simulated, FAO and NAT. A challenge however exists in validating the hindcast yields because of the "observed", i.e. official "observed" yield data are either the FAO statistics or nationally collected data based on methodologies that are unique to individual countries. In some instances, national (NAT) and FAO statistics did not match but the problem was solved by introducing a lag of one year to NAT yields. In some countries such as Tanzania, NAT and FAO yields differ remarkably while in some instances, available data were of short time series (10-years or less in Tanzania and Ethiopia). Yet again, subnational statistics from maize producing regions did not add up to the national reported yields. Further examination of the data sets did not reveal the sources of discrepancy and can be a point of investigation in further research. As an alternative to yield statistics, we might have used data from field trial stations but we did not have access to such. Therefore, to get a good time series for evaluating historical yield reforecasts, we had to resort to yield simulation driven by WFDEI.

We drove WOFOST with 15-ensemble seasonal climate forecasts resulting in 15-ensemble yield forecasts each corresponding to the climate forecasts. We compared the predicted average national yield anomalies to WFDEI yield anomalies to assess predictability of anomalies. National yields were aggregated based on planting dates and forecasts issued upto three months before planting dates (3-lead months). Standardized anomalies are used with the aim of detrending the time series. Prediction of annual yield anomalies depend on sowing dates and hence regions of the study area, for example predictability of maize yields planted in April in Kenya and Ethiopia are better than than planted in March and July. But some anomalies are potentially predicted in the entire region at least two months before planting. This was especially for years corresponding to extreme El-Niños i.e. 1984, 1989, 1997, 1998. In general, anomalous annual national yields are predictable at least 3-lead months before planting with the exception of Tanzania's.

This may be due to less variability in simulated yields in Tanzania in the study period i.e. mean deviation in yields over Tanzania if found to be low ($\approx$ 200Kg ha^{-1}).

To view spatial variability in predictability, we performed grid point validation (0.5°× 0.5° grid) based on sowing dates (i.e. all grid cells planted in March, April, July and December) and forecast lead-time. The AROC and ROCSS (see Appendix B.1) are used to assess probabilistic skill. More than 15% of grid cells planted in March (Kenya and Tanzania), April (Kenya and Ethiopia) and July (Ethiopia) predict above-normal and below-normal yields with significant ROCSS with three lead times. Near-normal forecast skill is lower but still predictable in the three lead time. This is not the same with yields planted in December (Tanzania) where forecast skill is lacking in all lead months. The proportion of grid cells with significant forecast skill of AN and BN yields change little with lead time up to lead-month two. A useful finding of yield forecast skill, with 2-months lead time before planting has the potential to change farmers' decisions with respect to crop, cultivar and other inputs choice so that the use of climate information is maximized. This may help reduce the gap between potential yields and actual farmers' yields especially in the current dispensation where actual yields in this region is found to be far below potential yields. Another measure used to evaluate probabilistic skill is the RPSS (Appendix B.2) which quantifies the closeness of cumulative probabilities of forecast to that of the observations. Yield show a significant skill in a larger percentage of cells in the region; probably because it only considers cumulative distribution (Mason, 2004) of the forecast rather than individual thresholds and cannot therefore reflect skill in a single tercile threshold when other terciles do not have skill. In this study therefore, the skill obtained from RPSS may not be useful as we required information on predictability of yield anomalies.

4. *Assess Influence of spatial aggregation from maize yield simulation grid on predictability*

   The unit of simulation in this research coincide with the FAO soil data at 0.1°×0.1° grids but can be aggregated to 0.5°×0.5° and further to sub-national (NUTS) boundaries. Simulated yields can also be aggregated by crop or similar soil types. This objective, (addressed in chapter 6) had three aims;

   - *to assess influence of yield aggregation on predictability by comparing the official observed agricultural statistics to simulated NUTS2 crop yields aggregated upwards from both half-degree and 0.1-degree resolutions*
   - *analyze and compare influence of aggregation on predictability based on 0.5 and 0.1-degree resolutions*
   - *asses predictability by aggregation over similar soils and varieties. In all cases we want to assess how the original grid before aggregation affects predictability of both yields and biomass (TAGP).*

To achieve the first aim we compared yields aggregated from 0.5°resolution (*NUT2_05*) and forecasts (*05FCST*); and from 0.1° to a sub-region (°NUT2_01) and forecasts (*FCST*) to case study regions of the study area. These were compared to observed data (*OBS*) to assess difference in average yields and variability. In North-Gonder, *NUT2_05* yields approximate the observed while NUT2_01 have significant difference to the observed by as much as 500Kg ha^{-1}. Variance in aggregated yields are quite low (CV=<10%) compared to the observed (CV=33%). Average lead-0 and lead-1 forecasts compare to *OBS*, but this forecasts aggregated from 0.1° resolution are quite high all lead-times assesed. The variance in all forecasts are quite low compared to *OBS*. Significantly lower yields than the observed is seen in Bungoma irrespective of aggregation. Forecasted yields are almost half the observed in all lead-times. Variability in OBS is higher i.e. CV =27%, 19% and 23% for *OBS*, *NUTS2* and *NUTS* aggregated yields respectively. The differences between *OBS* and simulated mean yields and their *CV* occur because some yield-limiting factors inherent in *OBS* are omitted in our model simulations as they are not incorporated in the WOFOST model. This includes factors such as effects of weeds and pests. Furthermore, water conserving techniques that are generally applied in the dryer regions are not considered. This may lead to an over estimation of dry spells stress. The fact that a simple bucket approach is used in our WOFOST version instead of a more detailed soil moisture module may contribute to this overestimation. The simple bucket soil moisture balance approach amplifies sensitivity to drought conditions in WOFOST. Fertilizer application is kept constant in our simulations but in reality, it may vary from one year or season to another hence influencing the observed yields. Less spread is obtained from *NUT2_01* forecasts than *NUT2_05*, the latter decreases with lead time because perhaps cells that are not harvested are eliminated before aggregation resulting in reduction of aggregated cells with lead time. At longer lead times, some cells do not grow to maturity yet the area cultivated remains the same. Thus this reduced yields when weighted by cultivated area. Difference may further be compounded with the fact that only one season is simulated in a single year yet some regions have double cropping seasons. This may be more true for Bungoma in which we see a consistent lower simulated yields irrespective of original grid before aggregation.

To achieve the second aim, we used aggregated yield forecasts from 0.5° resolution (*05FCST*) and from 0.1° resolution (*FCST* to assess influence of aggregation on predictability in the two study regions. In northern Ethiopia, we found that aggregation does not influence yield prediction since *FCST* and *05FCST* AN and BN tercile forecast skill (ROCSS≥0.3 and significant) are comparable. Below normal yield forecasts posses higher skill in all lead times. Aggregation however influence predictability of TAGP. Above normal *05FCST* forecasts of TAGP are predictable in all lead times (i.e lead 0-2) but only lead-0 *FCST* of the same tercile are predictable. In northern Ethiopia, there is no difference in yield prediction irrespective of the original grid before aggregation, but aggregation affects predictability of above ground biomass and shows less predictability from $FCST$ i.e. predictability of TAGP is a function of aggregation level and forecast tercile.

In equatorial Kenya, (Bungoma), we see the effects of aggregation on yields i.e. some predictability is seen when aggregated from 0.5° resolution i.e AN forecasts are predictable in all lead-times but not below normal forecasts while $FCST$ show no predictability in all terciles. TAGP is not predictable irrespective of original unit before aggregation. We have seen here that influence of aggregation is region, tercile, product (being assessed e.g. yield/biomass) dependent.

NUT regional dependency could result from the number of simulation units within each region. More simulation units may capture variability in crop model inputs such as cultivar properties, soils and weather characteristics that influence yields. For illustration, we see poor predictability in Bungoma, a region covered by nearly one 0.5° (WFDEI) (or 0.75° forecasts) climate grid. Variability in weather conditions and hence weather-climate-soil interactions at simulation grids are not well captured by WOFOST. It is known that rainfall has a higher spatial variability than temperature and it grossly affects rainfed yields. A larger area like North-Gonder, Ethiopia has several of the climate grids within its boundary and hence spatial variability and more details are better captured than in the former, resulting in better simulation. Aggregation over large areas like North-Gonder may improve yield prediction but it also averages spatial variability due to soil types and climate. meaning it may be good for yield (or biomass) estimation over a large region such as national or sub-national boundaries but may not suffice when more details are required, for example at farm scales. Good prediction when yields are aggregated in larger areas has been found in France (Supit and Van der Goot, 1999), in operational application in the then European CGMS (Boogaard et al., 2013). These simulate at higher resolution then aggregate output to larger areas. But aggregation of input soil data has also been found to affect yield simulation more than that of climate data (Ojeda et al., 2020). Decreasing the spatial extent of input data aggregation has also been found to improve simulations (Van Bussel et al., 2011) emphasizing the importance of high resolution crop modelling, even though the requirements for high resolution modelling differs from that of low resolution as highlighted in Ojeda et al. (2020). Yet in another study, aggregating crop model input fields has been found to introduce aggregation errors (Hoffmann et al., 2016) that were compounded in dry years and when soils are also aggregated. In our study, we aggregated yields to a lower resolution but carried out simulation at high resolution. Further research could also repeat a similar experiment with aggregated inputs to assess what works better. Users of skill when aggregated over large areas may differ from those who require high resolution products.

For example, early warning systems, relief agencies, governments may benefit from yield forecasts over large areas but not farmers, crop insurers and scientific applications.

The third aim was addressed by aggregating yields by crop variety over different soil types in the two sub-regions i.e. North-Gonder and Ethiopia. Yields are expected to vary with soil types due to differences in a number of soil parameters such as rooting depth, soil profiles, moisture holding capacity, soil fertility among others. Soil influences have been found in Olesen et al. (2000). In this study, we aggregated yields of crop varieties by soil type but found no significant variations in yield simulations/predictions by different soil types, irrespective of crop variety. This may be a result of almost similar available water capacities in the soils. Also, the maximum soil depth is fixed for all types in addition to the simple water balance in WOFOST. This does not allow influence of ground water in our simulation. Perhaps if we had more soil layers then root development would differ, and result in improved yield simulation and hence differences in predictability. For each region and crop variety, simulation and predictability of yields is better than for biomass perhaps because biomass production is sensitive to sunshine and radiation which do not vary much during a crops growing cycle in the region resulting in less predictability. In WOFOST, yield is a function of Harvest Index (HI), it is not constant but is influenced by climate during certain growth phases for example, rainfall during grain filling results in higher (HI) but will not have a similar effect on biomass. Harvest Index therefore varies annually and among the crop ensemble forecasts leading to variability in yields and possible prediction of anomalous events than in biomass.

7.2 Discussion on study Design, Data and Methods

This study was an exploration of the use of ensemble seasonal climate forecasts to predict yield in Eastern Africa based on the current tools and technologies available in both climate forecasting and crop modelling communities. The assumption here is that when we have good climate forecast (skillful) and use the same to drive an impact model, then the outcome should similarly have skill.This assumption may not be true for example, driving a crop model with seasonal climate forecasts may not guarantee skillful yield forecasts (Baigorria et al., 2007; Semenov and Doblas-Reyes, 2007; Shin et al., 2010). In some instances, impact models have been more skillful than the driving climate forecasts (McIntosh et al., 2005). Operationally, seasonal climate forecasts are routinely offered globally through regional climate outlook forums (COFs) such as the Greater Horn of Africa Climate Outlook Forum (GHACOF) (Ogallo et al., 2008) in the GHA. They develop consensus forecasts based on statistical methods, and dynamical ensemble climate forecasts from the global producing centers. We used one such ensemble prediction system system-4 (*S4*) from ECMWF (Molteni et al., 2011), one of the global forecast producing centers. We started by evaluating the skill of such a system using a range of forecast verification measures.

One challenge to forecast validation is the use of a reliable observed dataset. We therefore used the WFDEI with a resolution of 0.5°× 0.5° as the primary validation data set against *S4* at a resolution of 0.75°×0.75° . Other gridded data sets were also used to assess robustness of forecasts since the WFDEI and *S4* have similarities that would influence skill i.e. WFDEI is processed from ERA-Interim reanalysis and the same atmospheric model (IFS) used in production of daily weather forecasts in the reanalysis is also used in *S4*. Use of independent data sets was therefore a strength as it strengthens the robustness of results obtained.

Crop models have been used in agronomic research in many regions of the world and have been applied to assess climate impacts on crops. They have likewise been used for crop forecasting. This is the first study to use seasonal climate forecasts and crop models to forecast maize yields over a large area in eastern Africa. But crop models are of varied types that may not be universally applied across regions, we used one model from the Wageningen crop growth simulation models, WOFOST. Unless calibrated and validated in detail on data-rich experimental plots it is not easy to asses which model performs best in a given situation, or consequently to objectively select any particular crop model, or to use an ensemble of crop models. Use of ensemble of climate models to improve accuracy in simulation has been pioneered in the climate domain for example, the Climate Model Intercomparison Projects (Meehl et al., 2016). Similar to the climate domain, also in agronomy model intercomparison projects have been undertaken such as AgMIP (Rosenzweig et al., 2013). Again in analogy to climate models, Bassu et al. (2014) found that for maize crop models with low level calibration information, a single model may fail to accurately simulate absolute yield but that an ensemble of models is more likely to approach the correct absolute yield and better simulates variability. A similar setup with several models in this study could likely improve results.

Never the less, we used WOFOST. A model originally designed for a single location with homogeneous weather, soil and crop data; but in this study it was set up for a regional simulation meaning that input data within a single simulation unit may not be uniform. Properties occupying more that 50% of the simulation unit is assumed for the whole cell. Variability of weather may have been lost because of the coarse resolution of gridded data sets (≈0.5°) for the reference WFDEI. This implies that several cells of 10km would have similar weather characteristics. Because weather in this regional has high spatial variability, observed weather may not be representative though this may be different in the future as there currently exist very high resolution gridded observations for the region. For example, the 0.05° Climate Hazards group Infrared Precipitation with Stations (CHIRPS) data (Funk et al., 2015), and daily maximum and minimum temperature (CHIRTS-daily) data (Funk et al., 2019); the ECMWF ERA5 reanalysis (C3S, 2017; Hersbach, 2016); among others. Resolutions of forecasts are also improving and may be useful for similar future research with an assumption that high resolution climate data could capture well the spatial variability in weather. For example, the horizontal resolution of the atmosphere has increased to ≈36km in ECMWF's seasonal forecasting system (*SEAS5*) (Johnson et al., 2019) from ≈80km in *S4* climate forecasts (Molteni et al., 2011) in this study. At the start of this research, we used one of the best data sets available for verification and also climate forecasts from one of the world leading global forecast producing centers (Weisheimer and Palmer, 2014).

Results from this study must be used with an understanding of limitations which can be subject of further research. Many approaches have been used to estimate crop sowing dates in crop simulation experiments that include the use of observed planting dates, use of rainfall characteristics at the start of a season, and optimization of dates on maximum yields and minimum yield variance. Use of observed crop sowing dates was not available, and if it were, then matching the actual date and a regional crop model for WLY simulation would be problematic. This is because the choice of a farmers' planting dates are influenced by other factors such as availability of labor, seeds, farm inputs among others. Planting dates adopted in this study are optimized on maximum yields. The same dates are used every year of the study period hence no change in planting dates. This may not be true since dates most likely vary from year to year depending on rainfall and may thus result in yield losses in the model. The model is set to simulate only the first planting season in the study area, it therefore means that regions of the study area with two planting season would report lower yields. This may have contributed to the difference between WFDEI yields, national reported yields and FAO yield statistics. Though challenges facing agricultural statistics have noted (Keita and Carfagna, 2010). Hopefully, future research would have better observed yield data. However, because we were interested in prediction of yield anomalies, the method was just suitable. There exist a mismatch between climate forecast spatial resolution and crop model input requirements. This calls for post processing of climate forecasts by methods that are not limited to downscaling (whether statistical or dynamical methods), bias correction among others. The effect of bias correction on crop model simulation has not been assessed in this study. Also known is that GCM models are prone to producing many rainfall events and little rain per event (Hansen and Indeje, 2004), the influence of bias correction on these rain events was not explored.

7.3 Scientific contribution

Rainfall in the GHA is highly variable with frequent extremes and since agriculture is largely rain fed, climate change is expected to affect crop production. Current climate change could affect the mean climate, change the frequency of extremes, could change within season rainfall variability (i.e. rainy days, dry days, summer days, etc). These changes threaten crop production, but using seasonal climate forecast to predict yields may be one method of increasing farmers' resilience by enhancing the use of climate information. Farmers have always been adapting to climate variability, but current climate change poses newer threats because probable changes and impacts have not been observed before (Cairns et al., 2013). This study could make a contribution to establishing of a regional climate service for farmers, or a crop yield and performance monitor tailored towards the same principles as the European Unions' Crop Growth Monitoring System (CGMS) explained in Supit et al. (2012). Further assessment of skill could further our understanding of the robustness of the climate-impact modelling chain or forecasting across regions and lead-times

Predictability of maize yields with lead-times before planting dates would enable enable adjustment of farming practices before planting, but there also exist risks of unforeseen changes in weather during the growing season. We have found that while climate anomalies also drive yield anomalies over a large area such as the national level, and that localized sensitivity to yields depends on the variability or distribution of weather during a crops' growing cycle. Knowing weather characteristics that explain yields at each growth phase could be useful in tailoring targeted cheap interventions by scientists to conserve good yields in each locality without generalization of the crop management interventions in a large geographical area.

We assessed skill of both seasonal forecasts and yield forecasts over a large area of varying climatology, covering several agro-ecological zones using a crop model developed for field scale simulations. Such a design has enabled quick exploratory study without the need of field experiments and could be replicated elsewhere in the world. Many references in this manuscript have evaluated yield prediction skill focusing on several months lead time before harvesting. In this study we evaluate yield prediction prior to sowing. Many in the region have studied pre-season forecasting using statistical methods but not at a regional scale. To the best of our knowledge, this is a first attempt to perform such pre-season evaluation over a large area in eastern Africa using dynamic GCM climate forecast and a crop simulation model. Thus its added contribution to the body of knowledge in science.

7.4 Societal impacts and outlook

The Greater Horn of Africa is highly vulnerable to climate that often result in food insecurity. The region is characterized by repeated occurrences of drought and high variability in precipitation and a combination of other factors has reduced the ability of many farmers to respond before conditions deteriorate. This situation is gaining increasing attention of governments, development partners and civil organizations. Governments in particular have recognized such challenges and have developed policies and programs aimed at strengthening early warning systems, food security and agriculture.

Seasonal climate forecasts are operationally issued to users as a probability of excedence of a certain threshold (be it temperature or rainfall) or forecasts being within certain thresholds and accompanied by information on expected impacts on various sectors, including agriculture and water resources. Impacts outlook are normally based on expert opinions. This information must be understandable to the recipients if expected to influence any change in farm management activities, policy decisions either by relief agencies or governments to better manage associated climate risks (Troccoli et al., 2008; Jones et al., 2000). Forecasts should contain related uncertainty information. According to Jones et al. (2000), uncertainty limits the benefits of seasonal climate forecasts but Troccoli et al. (2008) notes that climate information and predictions play an essential role in management of risks associated with climate variability and change.It is hoped that future research would move this study to applications, working with farmers to assess performance of the system.

A number of early warning systems and organizations exist in East Africa (described in Chapter 5). These organizations provide their services largely to governments and humanitarian agencies and do not target the local farmers. For crop monitoring they use agrometeorological assessment reports and satellite technologies that monitor conditions of food crops after planting (e.g. NDVI) and rainfall to estimate impending food security situations. FEWSNET for example uses seasonal climate forecast in its assessments but does not provide explicit yield forecasts. This study can feed into the existing EWSs by providing pre-season yield forecasts to farmers in their locations. Based on yield forecasts, farmers may accordingly adjust their farm management practices. The forecasts could as well be used by, farm input commodity traders (for stocking of farm merchandise) besides extending the time horizons of the existing EWSs by providing forecasts of expected yields before planting. Crop insurance industry and future markets could also benefit from implementation of this study i.e. sign contracts for future crop supplies based on current prices.

Even though the results shown in this study are aggregated to half degree grid (weighted by planted area) and further to national country boundaries, the half degree provides a resolution presently lacking and may be useful to governments, insurance, and relief agencies because it captures the sub-national variability. The WOFOST crop model simulation in this study is done at FAO land use cells ($\approx$0.1-degree grid), with improved climate data grid resolution, a better product for farmers may be obtained.

The Climate Change Agriculture and Food Security (CCAFS) propose to develop an Integrated Agricultural Production and Food Security Forecasting System (INAPES) for East Africa (Kadi et al., 2011). Though initiatives like INAPES and the present thesis show promising potential, seasonal crop forecasting systems have not reached operational maturity yet. With such developments, this study is timely as it may provide useful insights. We have assessed seasonal climate skill for East Africa (Ogutu et al., 2017), and have gone ahead to demonstrate its usability in crop yield forecasting in this study. Our study can directly feed into the development of new systems envisaged for the future such as INAPES. Such developments could contribute towards efforts to sustainably intensify agricultural production, reduce poverty and enhance food security

Appendices

Appendix A

Quantile-quantile mapping (qqmap) bias correction method

The quantile mapping method transforms modelsimulated distribution through a transfer function with no priori assumption of the distribution. To implement this method for a variable X, ranked observed distribution is divided into a number of discrete quantiles. For each quantile division, a linear correction factor (transfer function) is calculated by dividing the mean observation in that quantile by the mean of the simulated variable in the same quantile and applied to simulated data Y, such that

$$Y_{t,i}^{corr} = \frac{1}{ecdf(X)_{t,i}}(ecdf(Y)_{t,j}^{raw}(X_{t,i})) \tag{A.1}$$

Where '*corr*' implies corrected model simulations while '*raw*' implies raw model output. Accuracy is controlled by the number of quantile divisions i.e. fewer quantiles might smooth out the information contained within the observed record while too many quantiles might result in over fitting of the data to the model (Jakob Themeßl et al., 2011; Lafon et al., 2013).

Appendix B

Probabilistic Validation

B.1 Relative operating characteristic curve (ROC) skill score (ROCSS)

For probabilistic forecasts, a warning is issued when the forecast probability for a predefined event exceeds some threshold. A set of hit-rates and false alarm rated are determined and plotted for these thresholds thus generating a ROC curve. The area under the ROC curve (AROC) indicates the performance of the forecast. Because there is skill only when the hit-rates are higher than the false alarm rates, the ROC curve lies above the 45° line for a skillful forecast (AROC > 0.5), and below (AROC $<$ 0.5) for a forecast that is no better than the climatology. AROC may be transformed into a skill $ROCSS = 2A - 1$ where $-1.0 \leq ROCSS \leq 1.0$. A score of 1.0 indicates a perfect forecast system; -1.0 indicates a perfectly useless forecast system and zero indicate no skill. The ROCSS expressed as a percentage quantifies the improvement over climatological forecast. For details, see Buizza and Palmer (1998); Mason and Graham (1999); Hamill and Juras (2006); Mason and Stephenson (2008); Barnston et al. (2010); Diro et al. (2012).

B.2 Ranked Probability Skill Score (RPSS)

RPS is a squared measure that compares the cumulative density function (CDF) of a probabilistic forecast with the CDF of the corresponding observation over a given number of discrete probability categories (Epstein, 1969; Weigel et al., 2007a). It measures the accuracy of multi-category probabilistic forecasts. Mathematically,

$$RPS = \frac{1}{K-1}\sum_{K-1}^{K}(CDF_{forecast,K} - CDF_{reference,K})^2 \quad \text{(B.1)}$$

In which $CDF_{forecast,K}$ and $CDF_{reference,K}$ are the predicted and observed probabilities for the kth forecast category (we use decile interval). To indicate the improvement of a multi-category probabilistic forecast relative to the reference forecast or climatology, this measure is converted to a skill score, the RPSS (Diro et al., 2012; Doblas-Reyes et al., 2003; Müller et al., 2005) measures the relative skill of the probabilistic forecast over that of the reference (climatology) in terms of getting close to the actual outcome.

$$RPSS = \frac{RPSS_{forecast} - RPS_{reference}}{RPS_{reference}} = 1 - \frac{RPS_{forecast}}{RPS_{reference}} \tag{B.2}$$

In which $RPS_{forecast}$ (expected value of RPS) is the squared difference between the forecast and the reference cumulative probabilities, and $RPS_{reference}$ is calculated by using the reference cumulative probability. A 100% RPSS implies a perfect probabilistic forecast while a negative value of RPSS means the skill of the forecast probability is worse than the reference. For details, see Barnston et al. (2010); Diro et al. (2012); Epstein (1969); Kumar et al. (2001); Manzanas et al. (2014); Mason (2004); Müller et al. (2005); Tippett and Barnston (2008); Weigel et al. (2007a,b).

Appendix C

Supplementary figures

C.1 Chapter-3 supplementary figures

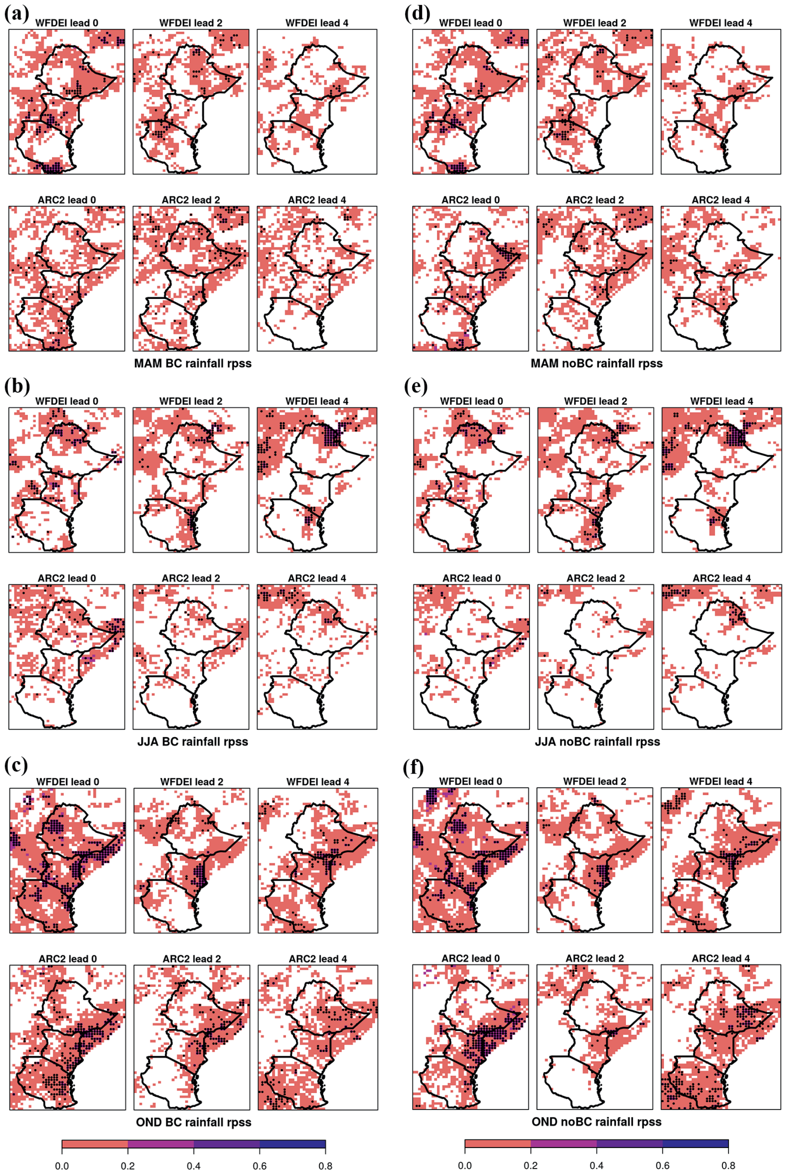


Figure C.1: Ranked probability skill score (RPSS) for bias-corrected [(a), (b), and (c)], and raw [(d), (e) and (f)] *S4* precipitation forecasts (*tp*) for MAM, JJA and OND seasons validated against WFDEI and ARC2. Only areas of skill (i.e. RPSS>0) are shown for lead months 0,2 and 4 before start of season.

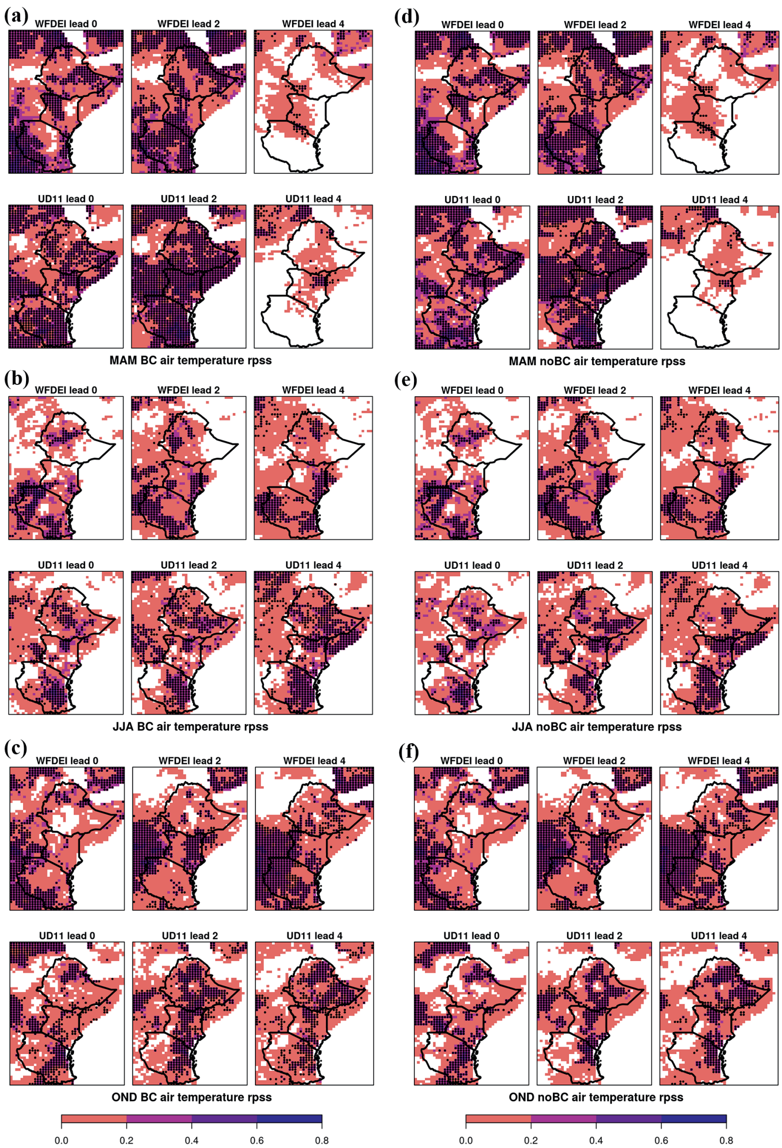


Figure C.2: Ranked probability skill score (RPSS) for bias-corrected [(a), (b), and (c)], and raw [(d), (e) and (f)] *S4* temperature forecasts (*tas*) for MAM, JJA and OND seasons validated against WFDEI and UD11. Only areas of skill (i.e. RPSS>0) are shown for lead months 0,2 and 4 before start of season.Dotted grids show areas where RPSS is significantly larger than zero at 95% significance level.

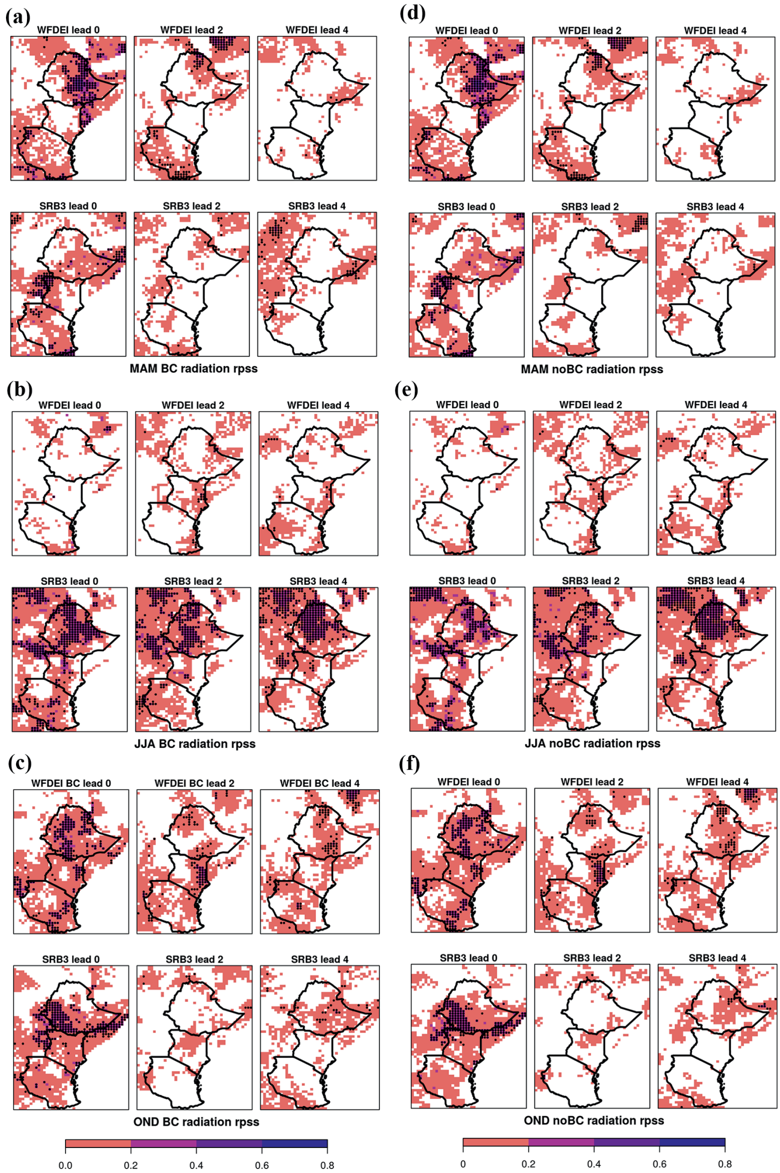


Figure C.3: Ranked probability skill score (RPSS) for bias-corrected [(a), (b), and (c)], and raw [(d), (e) and (f)] *S4* downward shortwave radiation forecasts (*rsds*) for MAM, JJA and OND seasons validated against WFDEI and SRB3. Only areas of skill (i.e. RPSS>0) are shown for lead months 0,2 and 4 before start of season.Dotted grids show areas where RPSS is significantly larger than zero at 95% significance level.

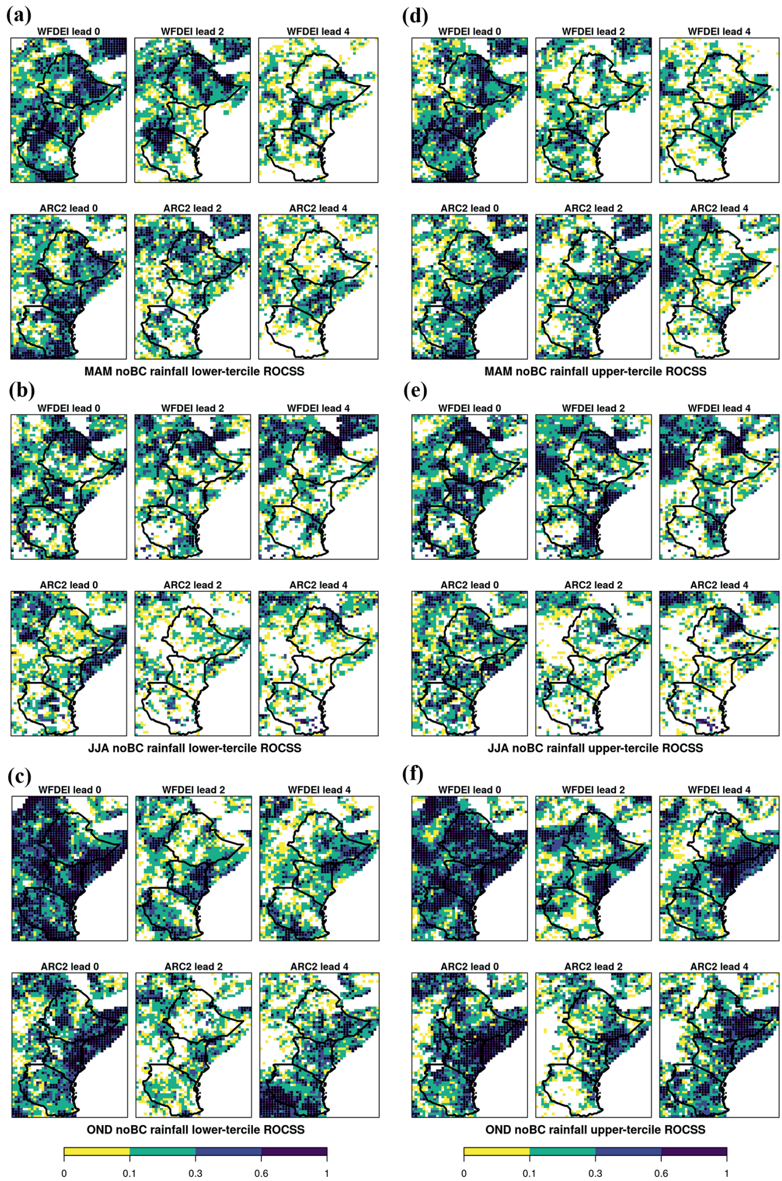


Figure C.4: Relative operating curve skill score (ROCSS) for lower-tercile (column 1), and upper-tercile (column 2) raw *S4* precipitation forecasts (*tp* for MAM (row 1), JJA (row 2) and OND (row 3) validated against WFDEI and ARC2. Only areas of skill (i.e. ROCSS>0) are shown for lead months 0, 2 and 4 before start of season. Dotted grids show areas where ROCSS is significantly larger than zero at 95% significance level.

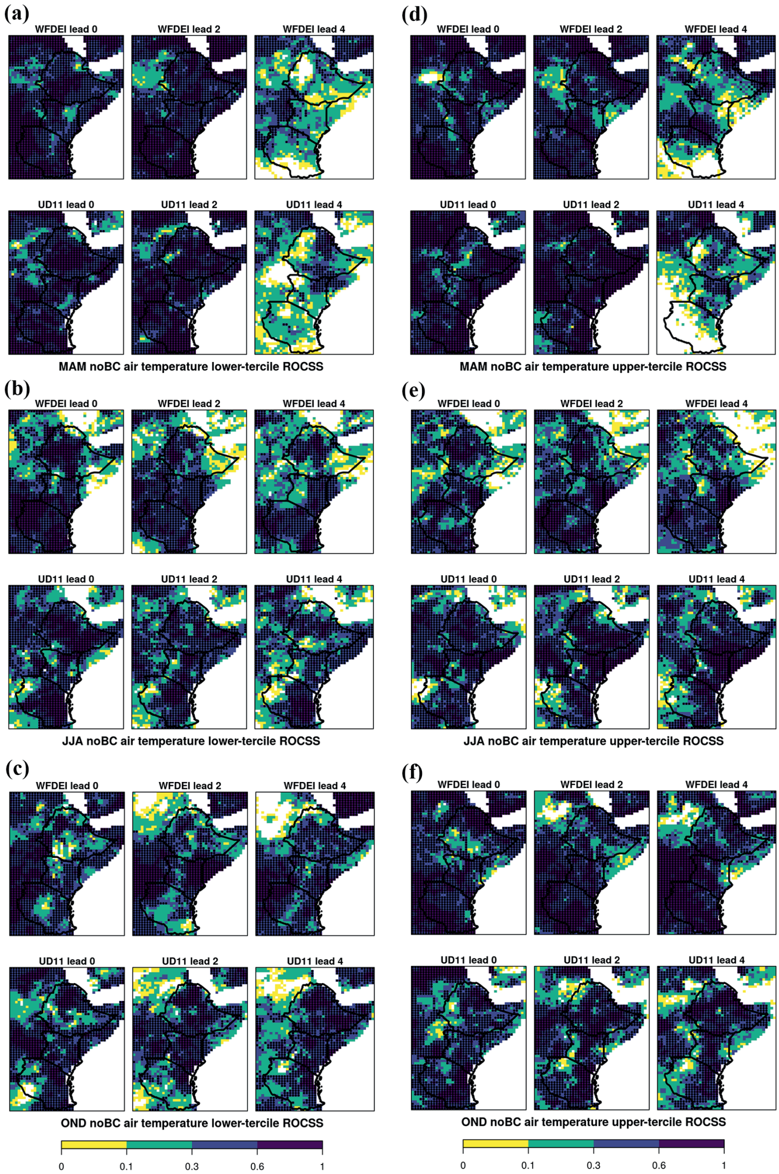


Figure C.5: Relative operating curve skill score (ROCSS) for lower-tercile (column 1), and upper-tercile (column 2) raw *S4* temperature forecasts (*tas* for MAM (row 1), JJA (row 2) and OND (row 3) validated against WFDEI and UD11. Only areas of skill (i.e. ROCSS>0) are shown for lead months 0, 2 and 4 before start of season. Dotted grids show areas where ROCSS is significantly larger than zero at 95% significance level.

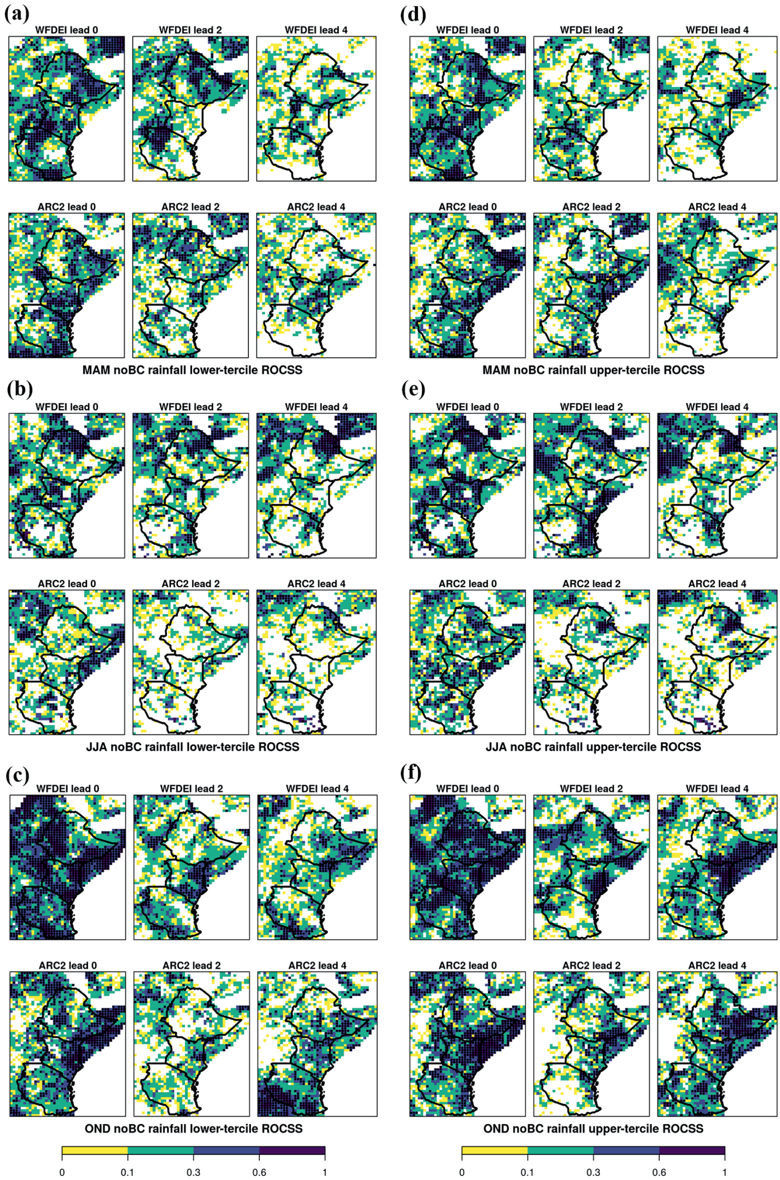


Figure C.6: Relative operating curve skill score (ROCSS) for lower-tercile (column 1), and upper-tercile (column 2) raw *S4* downward shortwave radiation forecasts (*rsds* for MAM (row 1), JJA (row 2) and OND (row 3) validated against WFDEI and ARC2. Only areas of skill (i.e. ROCSS>0) are shown for lead months 0, 2 and 4 before start of season. Dotted grids show areas where ROCSS is significantly larger than zero at 95% significance level.

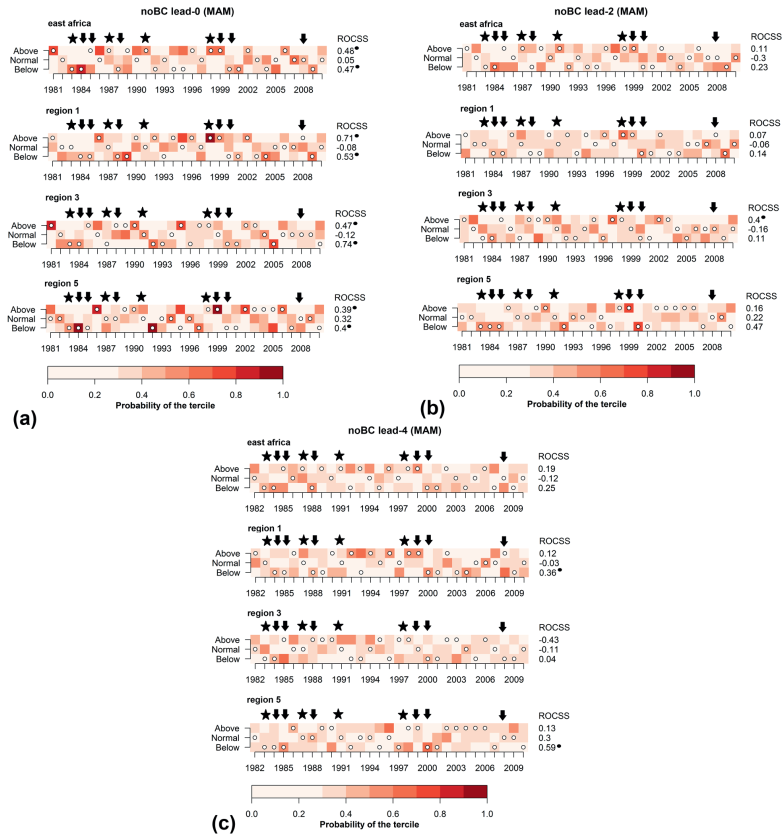


Figure C.7: Raw MAM season year-to-year precipitation forecast probabilities (colours), tercile of occurrence of observations (unfilled circles) and ROCSS for East Africa and sub-regions validated against WFDEI. Asterik indicate El-Niño years; arrows indicate La-Niña years and black dots indicate significant scores at 95% level.

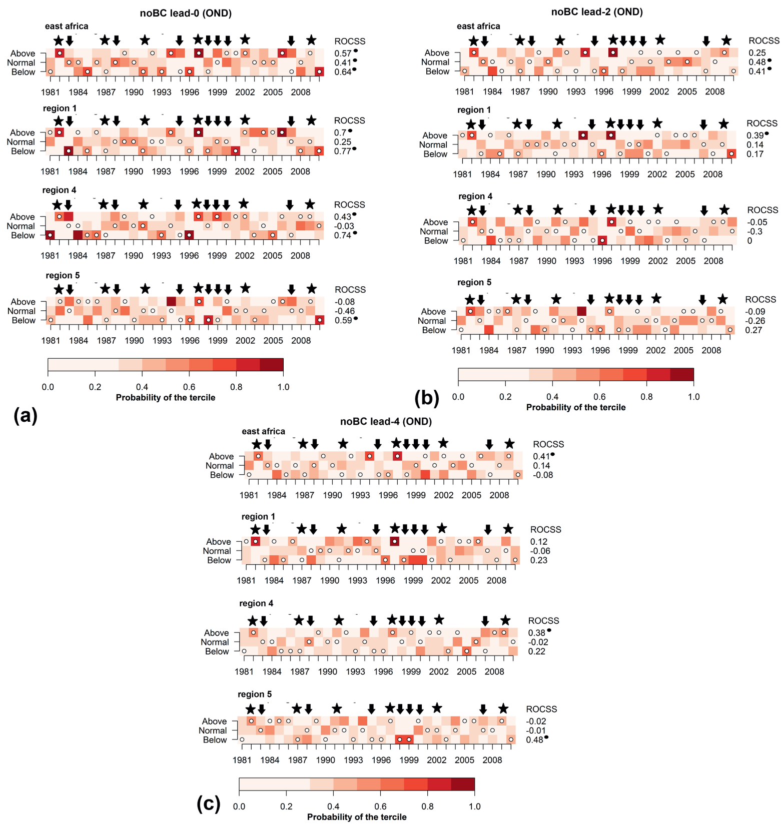


Figure C.8: Raw OND season year-to-year precipitation forecast probabilities (colours), tercile of occurrence of observations (unfilled circles) and ROCSS for East Africa and sub-regions validated against WFDEI. Asterik indicate El-Niño years; arrows indicate La-Niña years and black dots indicate significant scores at 95% level.

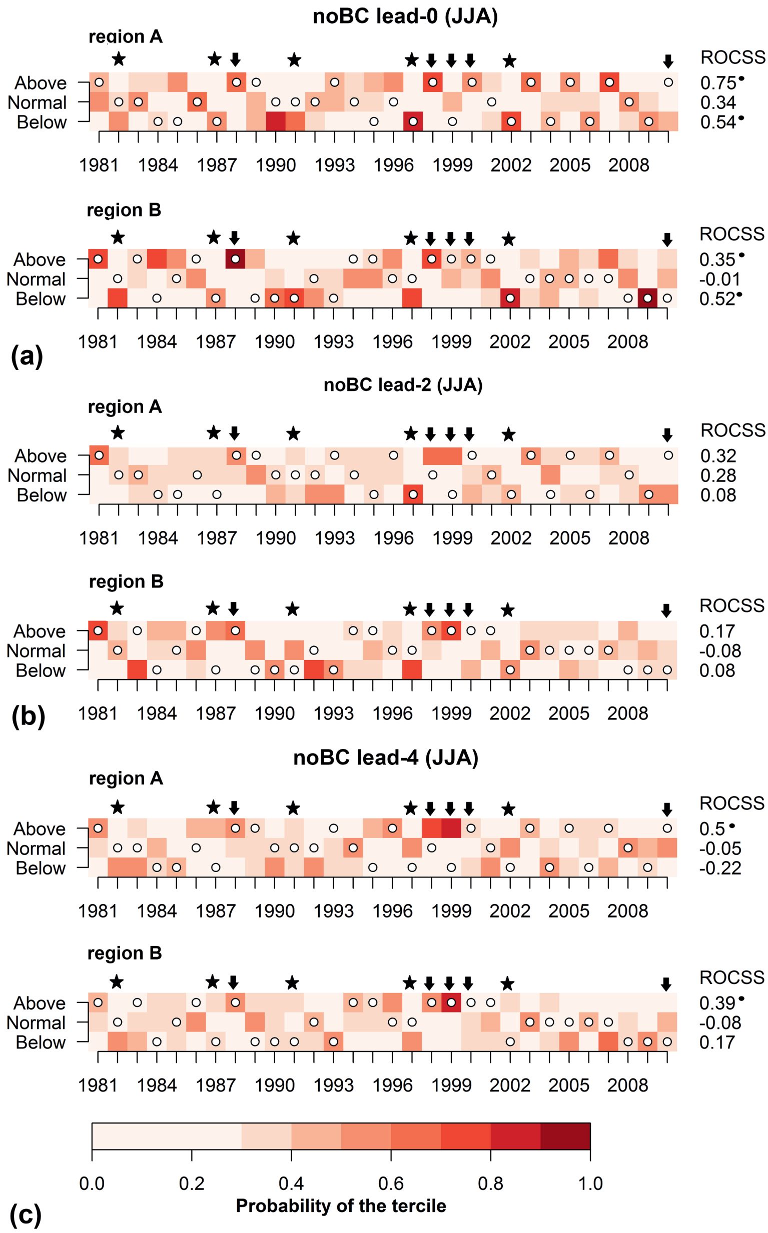


Figure C.9: Raw JJA season year-to-year precipitation forecast probabilities (colours), tercile of occurrence of observations (unfilled circles) and ROCSS for East Africa sub-regions in northern East Africa. Asterik indicate El-Niño years; arrows indicate La-Niña years and black dots indicate significant scores at 95% level.

C.2 Chapter-5 supplementary figures

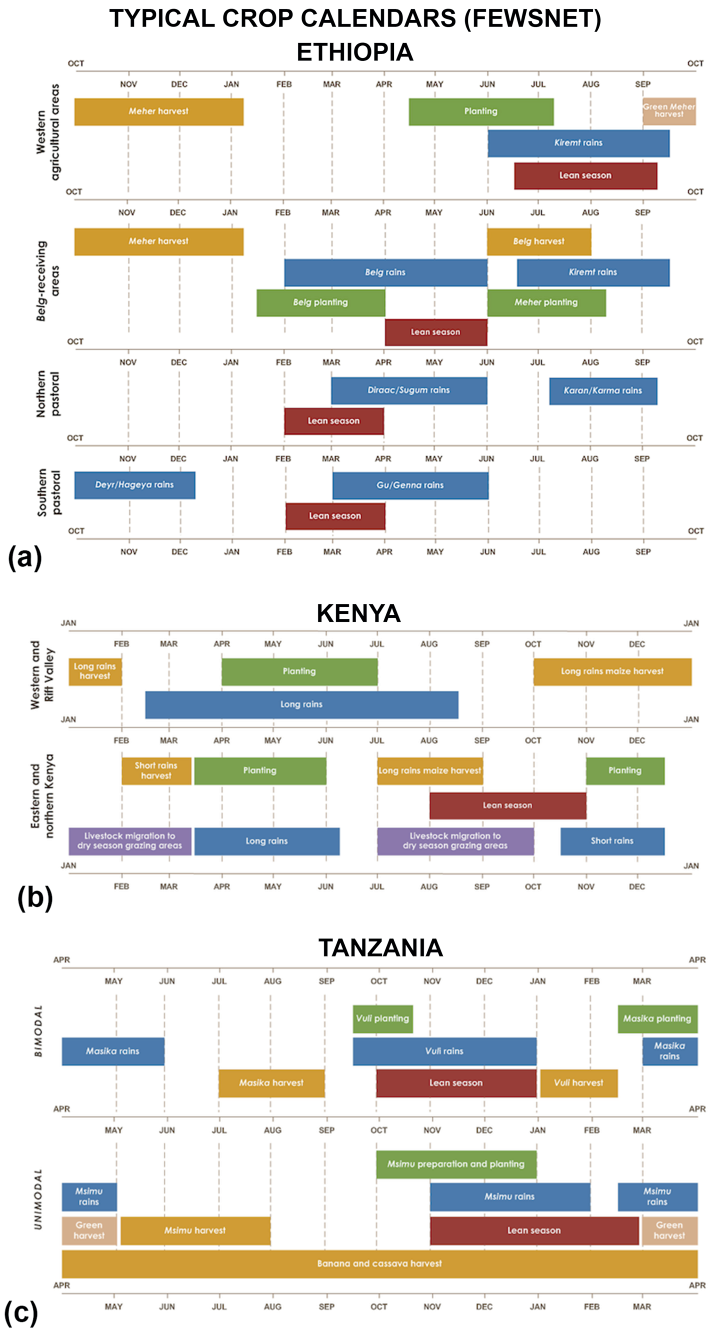


Figure C.10: Typical publicly available crop calendars, here from FEWSNET, for Ethiopia (a), Kenya (b), and Tanzania (c). These calendars are free available at https://fews.net/

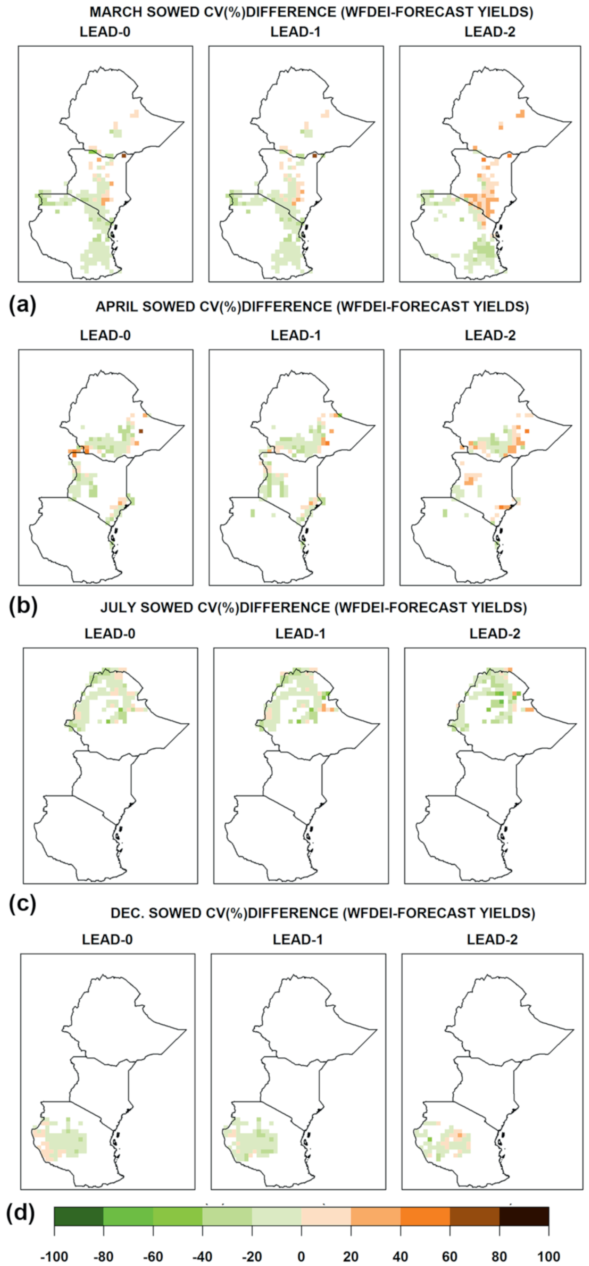


Figure C.11: Percentage difference in annual coefficient of variation (CV in percentage) between WFDEI and S4 yields for different sowing dates and forecast lead-times. Green (brown) colours imply that an S4 yield has a higher (lower) coefficient of variation than WFDEI yields

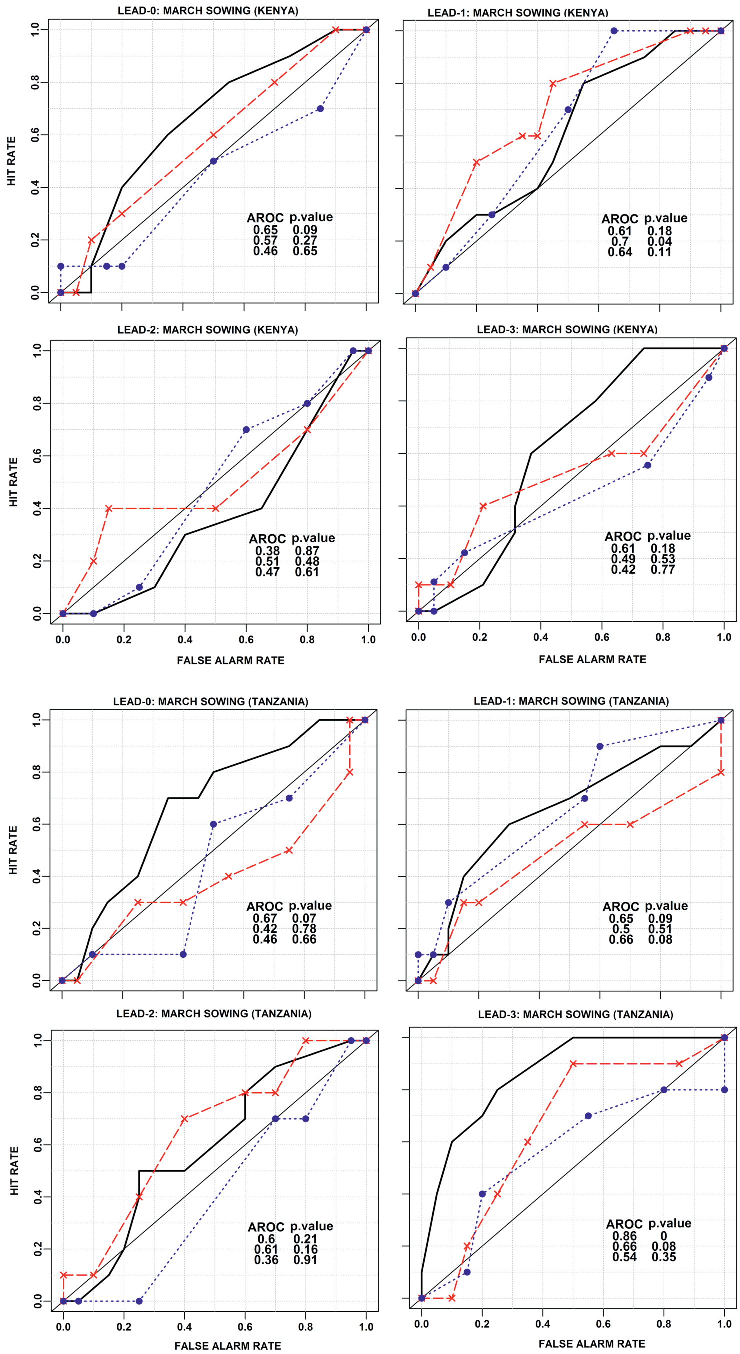


Figure C.12: Same as Figure 6 but for March sowing date in Kenya (a) and Tanzania.

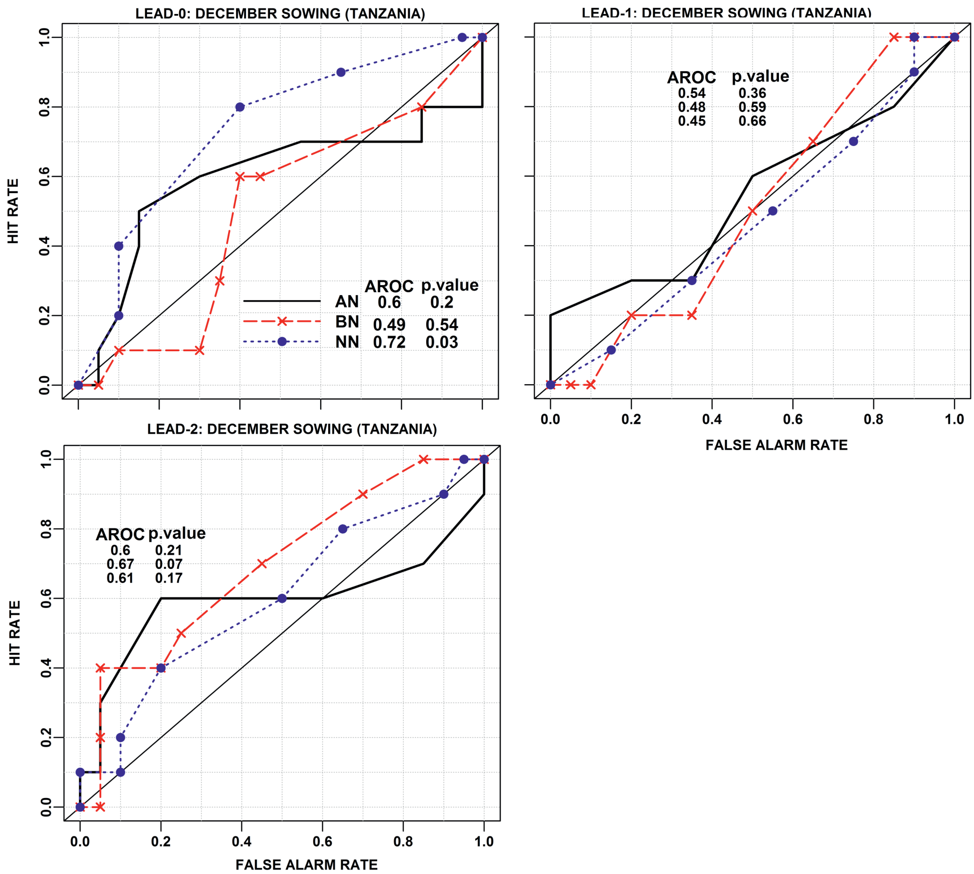


Figure C.13: Same as Figure 6 but for December sowing date in Tanzania.

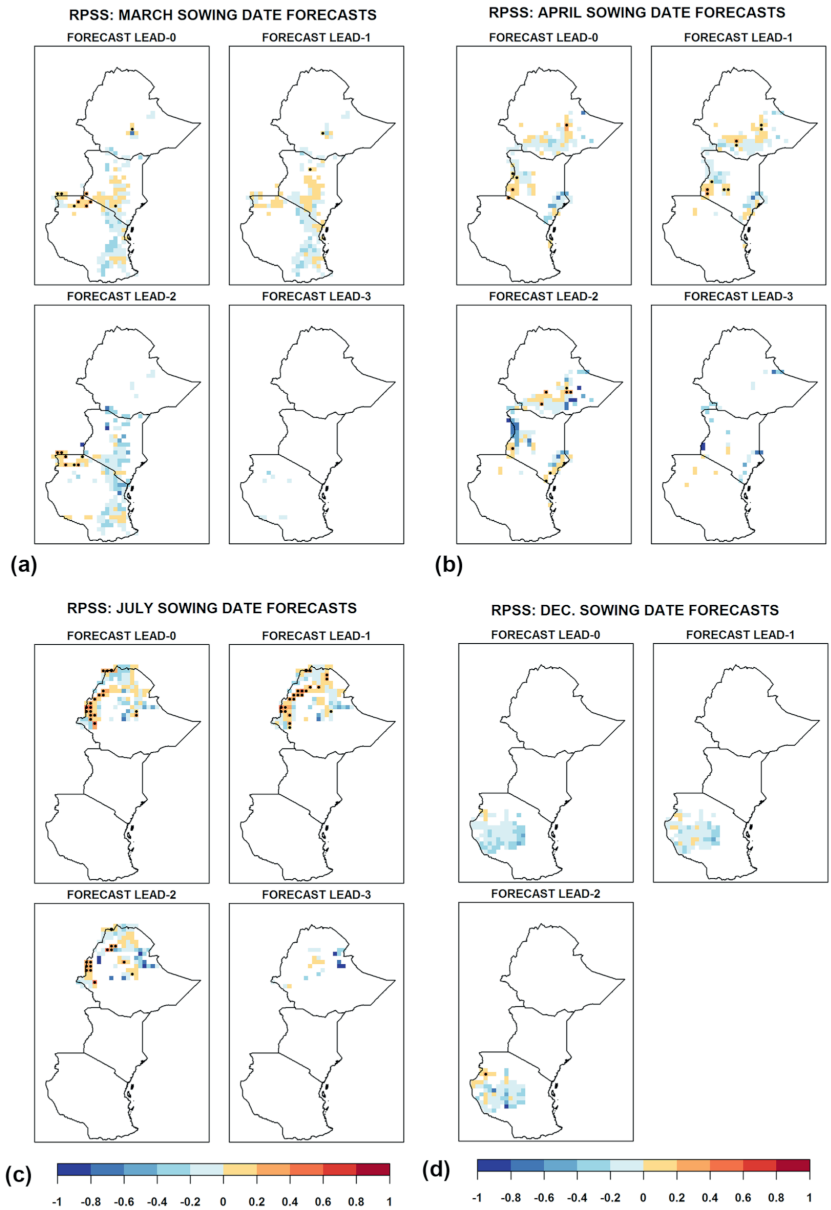


Figure C.14: Ranked probability skill score (RPSS) for yield forecasts for various harvest dates and forecast lead-times. Blue (red) colours show regions of no skill (skill). Dots indicate cells where RPSS is significantly greater than zero at 95% confidence level.

C.3 Chapter-6 supplementary figures

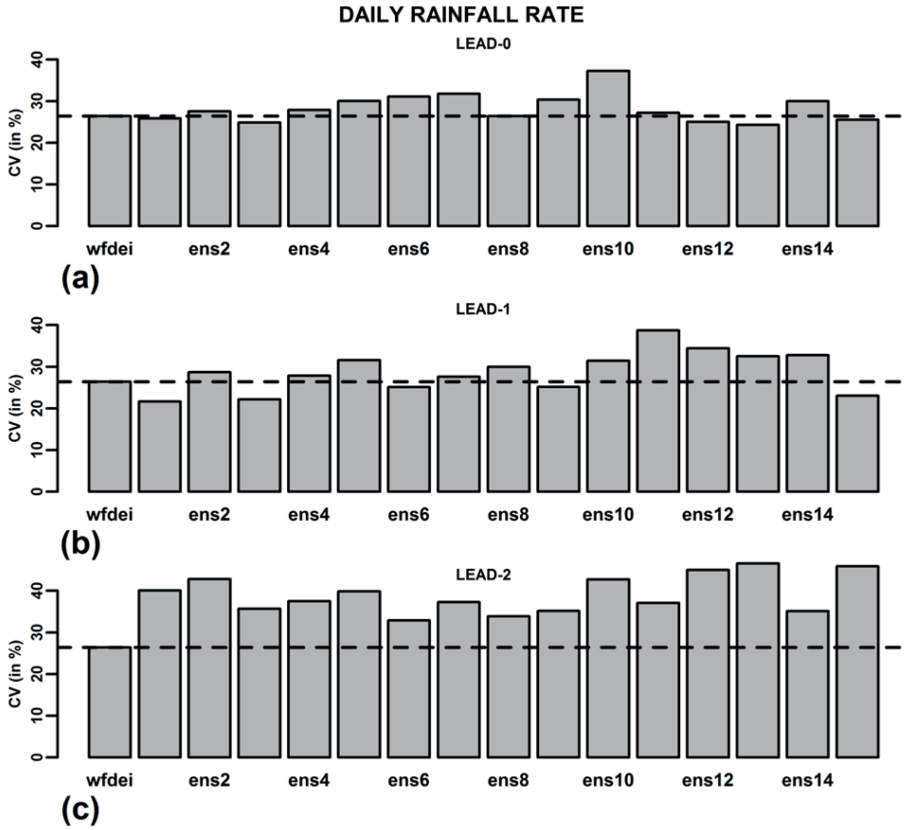


Figure C.15: Coefficient of variation (CV) of the reference precipitation data set (WFDEI) for Bungoma compared to that of ensemble members (ens1 to ens15). Horizontal dashed line shows CV of WFDEI for ease of comparison. CV is calculated for the period 1980-2010 and shown for three different lead times.

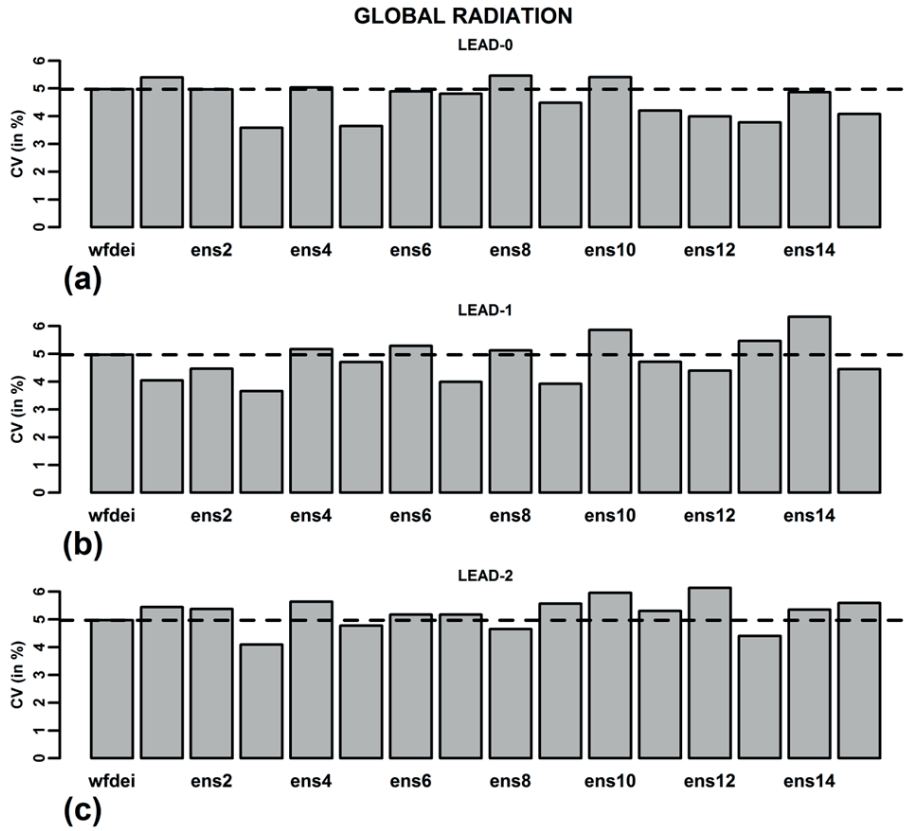


Figure C.16: Coefficient of variation (CV) of the reference global radiation data set (WFDEI) for Bungoma compared to that of ensemble members (ens1 to ens15). Horizontal dashed line shows CV of WFDEI for ease of comparison. CV is calculated for the period 1980-2010

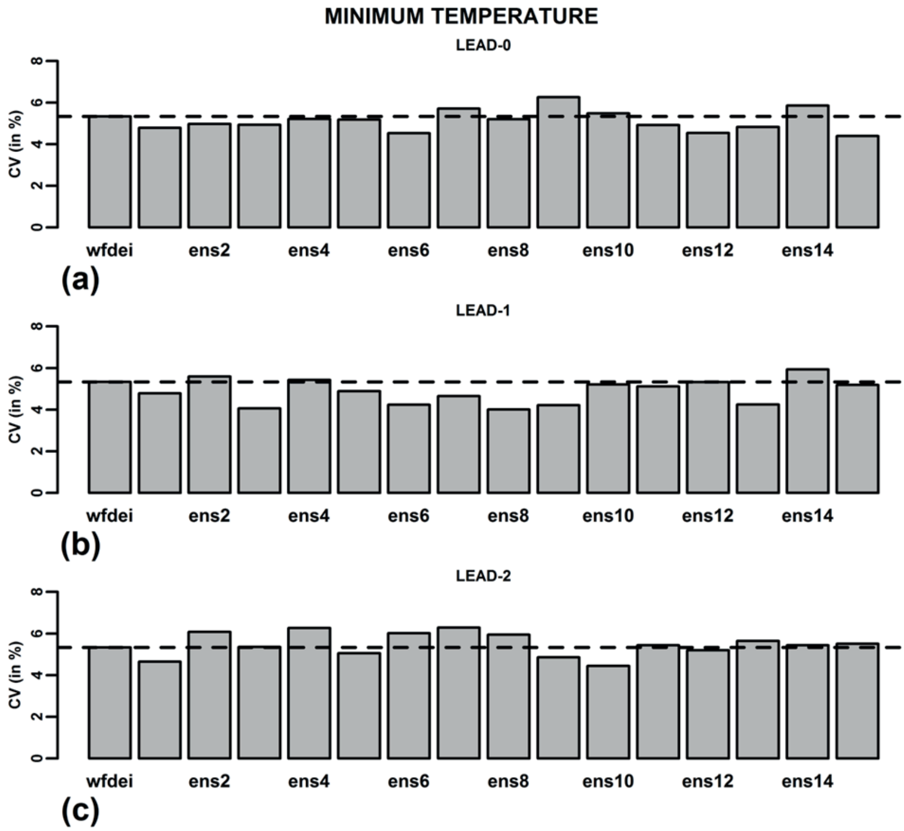


Figure C.17: Coefficient of variation (CV) of the reference minimum temperature data set (wfdei) for Bungoma compared to that of ensemble members (ens1 to ens15). Horizontal dashed line shows CV of wfdei for ease of comparison. CV is calculated for the period 1980-2010.

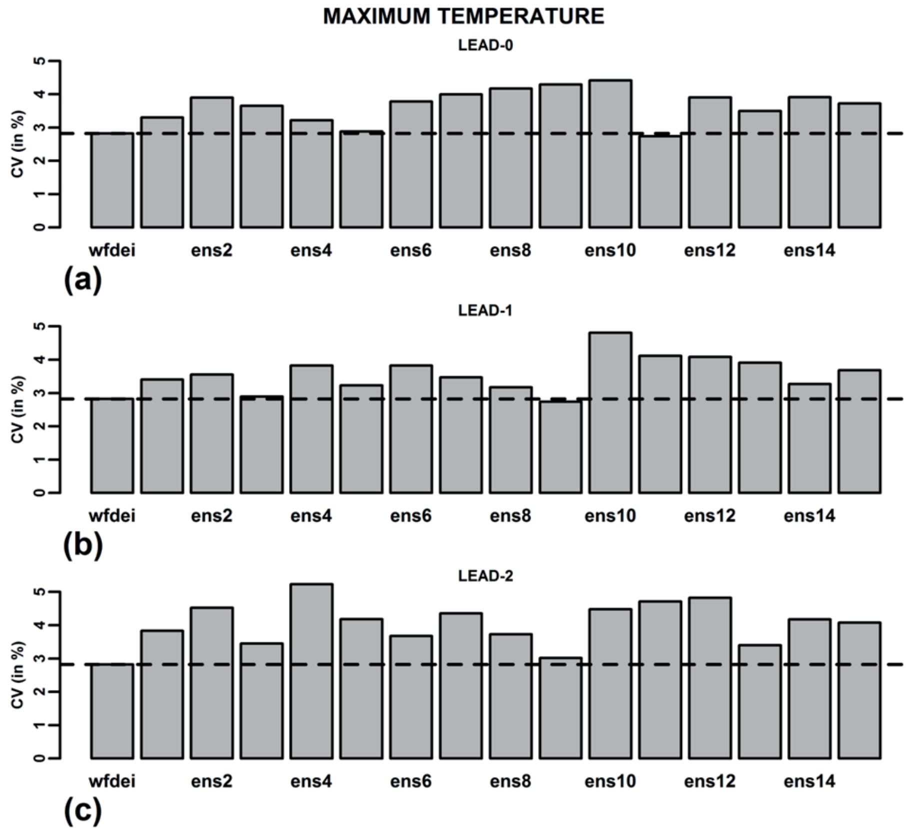


Figure C.18: Coefficient of variation (CV) of the reference maximum temperature data set (wfdei) for Bungoma compared to that of ensemble members (ens1 to ens15). Horizontal dashed line shows CV of wfdei for ease of comparison. CV is calculated for the period 1980-2010.

Appendix D

Supplementary tables

D.1 Chapter-4 supplementary tables

Table D.1: Forecast verification of North-Gonder growing season climate indicators

	indicator	SprErr	GDS(%)	r	bias	pbias(%)	MAE	RMSE	meanfcst	meanobs
	RR	0.94	57	0.2	0.4	7.9	0.7	0.9	5	4.6
	UR	0.91	54	0.2	-7.5	-5.1	29.2	37.4	138.8	146.2
	ER	0.95	58	0.2	0.4	7.7	0.7	0.9	4.9	4.6
	Ad	0.91	54	0.1	-1654.4	-9.4	4206.4	5400.5	15976.4	17630.8
lead-0	TX	0.72	63	0.4	-0.5	-1.8	0.7	0.9	26.5	27
	TN	0.63	62	0.3	-0.2	-1.2	0.5	0.5	14.4	14.6
	TG	0.6	70	0.5	-0.3	-1.6	0.4	0.6	20.5	20.8
	GDD	0.6	45	-0.2	-154.7	-10.4	156.5	268.2	1339.6	1494.3
	KDD	1.54	69	0.6	-0.7	-10.4	4.2	5.7	5.7	6.4
	RR	0.84	58	0.2	0.3	3.7	0.6	0.9	4.9	4.6
	UR	0.9	65	0.5	-1.1	-0.7	25	31.1	145.1	146.2
	ER	0.85	60	0.3	0.3	6.5	0.6	0.9	4.9	4.6
	Ad	0.97	61	0.2	-1576.9	-8.9	3910.4	5103	16053.9	17630.8
lead-1	TX	0.8	61	0.3	-0.4	-1.6	0.7	0.9	26.6	27
	TN	0.66	59	0.3	-0.2	-1.2	0.5	0.5	14.4	14.6
	TG	0.7	67	0.5	-0.3	-1.4	0.4	0.6	20.5	20.8
	GDD	0.4	46	-0.2	-123.8	-8.3	128.7	236.9	1370.5	1494.3
	KDD	2.1	66	0.8	0.02	0.3	3.3	4.5	6.4	6.4
	RR	0.74	65	0.4	0.5	9.8	0.7	0.9	5.1	4.6
	UR	0.87	58	0.4	-1.1	-0.8	26.3	32.9	145.1	146.2
	ER	0.74	66	0.4	0.5	9.8	0.7	0.8	5	4.6
	Ad	0.92	53	0.1	-1471.5	-8.3	4138.7	5359	16159.3	17630.8
lead-2	TX	1	63	0.3	-0.5	-1.7	0.7	0.8	26.6	27
	TN	0.7	64	0.4	-0.1	-0.7	0.4	0.5	14.5	14.6
	TG	0.9	67	0.5	-0.3	-1.3	0.4	0.5	20.5	20.8
	GDD	0.14	40	-0.3	121	-8.1	121	200	1373.3	1494.3
	KDD	1.8	63	0.5	-0.2	-3.5	4.7	6.5	6.2	6.4

Table D.2: Forecast verification of North-Gonder vegetative stage climate indicators

	indicators	SprErr	GDS(%)	r	bias	pbias(%)	MAE	RMSE	meanfcst	meanobs
lead-0	RR	0.69	71	0.5	0.1	1	0.8	1.2	7.5	7.5
	UR	1	52	0.2	3.1	8	14.4	16.3	45.1	41.2
	ER	0.68	71	0.5	0.1	1	0.8	1.2	7.4	7.3
	Ad	0.93	53	0.2	102.1	4	1107.5	1299	2730.8	2628.7
	TX	0.59	61	0.3	-0.4	-2	0.7	0.9	25.7	26.1
	TN	0.47	61	0.4	-0.4	-2	0.6	0.7	15	15.3
	TG	0.5	64	0.3	-0.4	-2	0.5	0.7	20.3	20.7
	GDD	0.13	39	-0.2	-62.1	-8	64.2	178	759.8	822
	KDD	1.04	61	0.2	-0.2	-7.90	2.1	3.1	2	2.2
lead-1	RR	0.72	71	0.5	0.2	2	0.8	1.2	7.6	7.5
	UR	1.1	55	0.1	3.6	9	13.7	16.5	45.6	42
	ER	0.7	71	0.5	0.2	2	0.8	1.2	7.5	7.3
	Ad	1.1	57	0.2	96.8	4	1019.5	1284.7	2725.4	2628.7
	TX	0.68	57	0.2	-0.3	-1	0.7	0.9	25.7	26.1
	TN	0.58	62	0.4	-0.3	-2	0.5	0.6	15	15.3
	TG	0.57	61	0.3	-0.3	-2	0.5	0.7	20.4	20.7
	GDD	0.13	33	-0.3	-54.4	-7	59.3	177.2	767.6	822
	KDD	1.27	59	0.2	0.5	24	2.4	3.6	2.7	2.2
lead-2	RR	0.7	68	0.4	0.5	7	1	1.3	7.9	4.5
	UR	1.1	61	0.3	-3.7	-9	12.3	14.7	38.2	42
	ER	0.7	68	0.4	0.5	7	1	1.3	7.8	7.3
	Ad	1.1	58	0.2	-740.2	-28	1093.4	1430	1888.5	2628.7
	TX	0.92	61	0.2	-0.4	-2	0.7	0.9	25.7	26.1
	TN	0.58	62	0.4	-0.3	-2	0.5	0.6	15	15.3
	TG	0.8	63	0.3	-0.3	-2	0.5	0.7	20.4	20.7
	GDD	0.26	42	-0.2	-54.9	-7	61.2	176.1	767	822
	KDD	1.85	62	0.1	0.3	14	2.3	3.4	2.5	2.2

Table D.3: Forecast verification of North-Gonder climate indicators around anthesis

	indicator	SprErr	GDS(%)	r	bias	pbias	MAE	RMSE	meanfcst	meanobs
lead-0	RR	1.2	54	0.1	0.2	5.1	0.8	1	3.1	3
	UR	0.4	61	0.1	0.1	0.9	3.6	4.5	10	9.9
	ER	0.2	54	0.1	0.1	5.1	0.8	1	2.9	2.8
	Ad	1.2	54	-0.02	-24.8	-15.3	96.8	131	137	161.7
	TX	0.2	53	0.2	-0.5	-1.7	0.7	0.9	27.2	27.6
	TN	0.9	64	0.4	-0.1	-1.0	0.5	0.7	14.3	14.4
	TG	1	65	0.4	-0.3	-1.4	0.5	0.6	20.7	21
	GDD	0.5	66	0.1	-30.9	-12.1	36.6	103.8	225.8	256.7
	KDD	1.3	48	0.1	-0.2	-11.5	1.7	2.4	1.4	1.6
lead-1	RR	1.3	54	0.1	0.2	6.1	0.8	1	3.2	3
	UR	1.25	50	-0.04	-0.1	-0.7	3.6	4.7	9.8	9.9
	ER	1.24	54	0.1	0.2	5.8	0.8	1	3	2.8
	Ad	1.16	43	-0.1	-19.4	-12.0	98.7	132.7	142.3	161.7
	TX	0.3	50	0.1	-0.4	-1.3	0.7	0.9	27.2	27.6
	TN	0.9	65	0.4	-0.2	-1.3	0.5	0.6	14.3	14.4
	TG	1	67	0.5	-0.3	-1.4	0.4	0.6	20.7	21
	GDD	0.1	67	0.2	-31.9	-12.4	34.8	103.8	225	256.7
	KDD	1.3	50	0.2	-0.1	-8.9	1.5	2.2	1.5	1.6
lead-2	RR	1.1	59	0.2	0.7	22.6	0.9	1.2	3.7	3
	UR	1.23	57	0.1	0.3	3.4	3.6	4.6	10.1	9.9
	ER	1.05	58	0.2	0.6	22.5	0.8	1.1	3.4	2.8
	Ad	1.25	58	0.2	-4.4	-2.7	84.4	117	157.3	161.7
	TX	0.94	45	-0.2	-0.8	-2.9	0.9	1.13	26.8	27.6
	TN	0.9	65	0.5	-0.1	-0.4	0.5	0.6	14.4	14.4
	TG	0.9	61	0.3	-0.4	-2.1	0.6	0.7	20.6	21
	GDD	0.1	59	0.1	-34.8	-13.6	37.4	105.3	221.8	256.7
	KDD	1	42	-0.1	-0.7	-42.4	1.5	2.4	0.9	1.6

Table D.4: Forecast verification for North-Gonder reproductive stage climate characteristics

	indicator	SprErr	GDS(%)	r	bias	pbias(%)	MAE	RMSE	meanfcst	meanobs
lead-0	RR	1.43	40	-0.2	0.2	13.7	0.6	0.8	1.5	1.3
	UR	1	40	-0.2	0.4	2	8.3	10.9	18.1	17.8
	ER	1.5	40	-0.2	0.1	11.8	0.6	0.8	1.4	1.2
	Ad	1.1	44	-0.1	0.9	0.1	521.9	658.9	868.7	867.8
	TX	1.3	66	0.5	-0.4	-1.30	0.6	0.7	27.8	28.2
	TN	1	57	0.2	0.02	0.20	0.5	0.6	13.6	13.6
	TG	1.1	70	0.6	-0.2	-0.90	0.4	0.5	20.7	20.9
	GDD	1	64	0.3	-18.8	-2.90	53.7	93.3	637.8	656.6
	KDD	0.5	66	0.6	-0.5	-10.70	3.2	4.6	3.8	4.3
lead-1	RR	1.2	59	0.4	-0.2	-12.00	0.5	0.6	1.1	1.3
	UR	1.4	53	0.3	3	16.70	8	9.5	20.8	17.8
	ER	1.1	60	0.4	-0.2	-12.30	0.5	0.6	1.1	1.2
	Ad	1.5	62	0.4	250.4	28.90	472.9	572.9	1118.2	867.8
	TX	1.2	60	0.4	-0.3	-1.10	0.6	0.7	27.8	28.2
	TN	0.8	58	0.2	-0.2	-1.50	0.6	0.7	13.4	13.6
	TG	1	68	0.5	-0.3	-1.30	0.4	0.6	20.6	20.9
	GDD	0.4	54	0.1	39.1	6.00	48	98.6	696	656.6
	KDD	2.3	64	0.9	-0.4	-10.40	2.3	3	3.8	4.3
lead-2	RR	1.3	56	0.1	0.1	11.50	0.5	0.6	1.4	1.3
	UR	1	42	-0.1	4.1	23.00	9	11.4	21.9	17.8
	ER	1.2	56	0.1	0.1	10.90	0.5	0.6	1.4	1.2
	Ad	1.1	46	0.1	197.5	22.80	511.6	651.1	1065.3	867.8
	TX	1.3	62	0.3	-0.4	-1.50	0.7	0.8	27.7	28.2
	TN	0.9	64	0.4	-0.01	-0.10	0.4	0.6	13.6	13.6
	TG	1.2	67	0.5	-0.2	-1.10	0.4	0.5	20.7	20.9
	GDD	0.5	61	0.1	-17.7	-2.70	51.5	94.5	638.9	656.6
	KDD	1.6	59	0.6	-0.7	-16.00	3.3	4.8	3.6	4.3

Table D.5: Forecast verification of Bungoma-Kenya growing season climate indicators

	indicator	SprErr	GDS(%)	r	bias	pbias	MAE	RMSE	meanfcst	meanobs
lead-0	RR	0.7	55	0.2	-0.2	-5.1	0.7	1	4.2	4.5
	UR	0.5	41	-0.2	-7.6	-16.8	19.6	25.4	37.9	45.5
	ER	0.7	54	0.2	-0.2	-5.1	0.7	1	4.2	4.4
	Ad	0.6	51	0.1	-421	-16.7	4764.1	6080.5	2100.2	2521.3
	TX	1	68	0.5	-0.05	-0.2	0.4	0.5	26.7	26.8
	TN	0.5	66	0.4	-0.2	-1.2	0.5	0.6	13.5	13.7
	TG	0.8	70	0.6	-0.1	-0.5	0.4	0.5	20.1	20.2
	GDD	1.55	54	0.1	-8.2	-0.5	10.9	14.8	1556	1564.2
	KDD	1.1	69	0.4	1.1	5.5	10.1	12.8	21.2	20.11
lead-1	RR	0.7	52	0.1	-0.1	-2.2	0.7	1	4.4	4.5
	UR	0.7	43	-0.04	-6.3	-13.8	17.8	23.7	39.2	45.5
	ER	0.7	52	0.1	-0.1	-2.2	0.7	1	4.3	4.4
	Ad	0.7	48	-0.1	-389.4	-15.4	5010.4	6317.5	2131.9	2521.3
	TX	0.3	76	0.7	-0.04	-0.1	0.4	0.5	26.7	26.8
	TN	0.6	67	0.5	-0.1	-0.9	0.5	0.6	13.6	13.7
	TG	1	73	0.7	-0.1	-0.4	0.3	0.4	20.1	20.2
	GDD	1.3	66	0.2	-6.7	-0.4	9.1	10.9	1557.5	1564.2
	KDD	1.3	63	0.3	1.8	9.1	10.4	12.5	21.9	20.1

Table D.6: Forecast verification of Bungoma-Kenya vegetative stage climate indicators

	indicator	SprErr	GDS(%)	r	bias	pbias	MAE	RMSE	meanfcst	meanobs
lead-0	RR	0.7	53	0.1	-0.3	-6.50	0.8	1.1	4.2	4.5
	UR	1	65	0.4	5.4	16.50	13.5	15.9	38.1	32.7
	ER	0.7	52	0.1	-0.3	-6.80	0.8	1.1	4.1	4.4
	Ad	0.8	65	0.4	710.4	38.90	1418.8	1841.9	2537.1	1826.6
	TX	1	65	0.4	0.04	0.20	0.5	0.6	26.4	26.36
	TN	0.5	65	0.4	-0.2	-1.50	0.5	0.6	13.7	13.9
	TG	0.7	71	0.6	-0.1	-0.70	0.4	0.5	20.1	20.2
	GDD	1	51	0.1	-2.7	-0.30	4.7	6.9	841.5	844.2
	KDD	1.1	66	0.4	-0.3	-3.10	5.2	6.4	8	8.2
lead-1	RR	0.8	55	0.03	-0.1	-3.00	0.8	1.1	4.4	4.5
	UR	1	53	0	5	15.40	15.3	17.9	31.7	32.7
	ER	0.8	54	0.02	-0.1	-3.10	0.8	1.1	4.2	4.4
	Ad	1	53	0.1	535.1	29.30	1619.8	1939.8	2361.7	1826.6
	TX	1.5	70	0.6	0.1	0.30	0.4	0.5	26.4	26.36
	TN	0.6	66	0.5	-0.2	-1.10	0.5	0.6	13.8	13.9
	TG	1	73	0.7	-0.1	-0.50	0.3	0.5	20.1	20.2
	GDD	1	51	0.02	-3.1	-0.40	5	7.2	841.1	844.2
	KDD	1.5	66	0.2	0.7	7.90	5.5	6.8	8.9	8.2
lead-2	RR	0.9	55	0.2	-0.1	-2.70	0.8	1	4.4	4.5
	UR	1.1	55	0.1	5.6	17.00	15.9	17.6	38.2	32.7
	ER	0.8	54	0.2	-0.1	-2.90	0.8	1	4.3	4.4
	Ad	1	45	-0.1	653.9	35.80	1685	2067.6	2480.5	1826.6
	TX	1.2	66	0.4	0.1	0.20	0.5	0.6	26.4	26.36
	TN	0.7	70	0.5	-0.1	-1.00	0.5	0.6	13.8	13.9
	TG	1	71	0.6	-0.1	-0.50	0.4	0.5	20.1	20.2
	GDD	1	59	0.1	-3	-0.40	4.8	7.2	841.2	844.2
	KDD	1.7	61	0.3	0.2	2.20	5.3	6.5	8.4	8.2

Table D.7: Forecast verification for Bungoma, Kenya anthesis stage climate indicators

	Indicator	SprErr	GDS(%)	r	bias	pbias(%)	MAE	RMSE	meanfcst	meanobs
lead-0	RR	1	61	0.2	-1	-25.6	1.2	1.8	2	4.1
	UR	1	59	0.2	-0.5	-5.2	4.5	7	10	10.5
	ER	0.9	59	0.2	-1	-25.9	1.2	1.8	2.9	3.9
	Ad	0.9	50	0.1	21.3	-98.2	166.1	221.2	-0.4	-21.7
	TX	1.2	57	0.2	0.4	1.4	0.8	1	26.7	26.3
	TN	1	67	0.4	-0.1	-0.6	0.5	0.7	13.56	13.64
	TG	1.1	61	0.3	0.2	0.7	0.5	0.7	20.1	20
	GDD	1.1	60	0.3	2.7	1.3	10.6	13.8	211.8	209
	KDD	1.8	53	0.2	0.9	149.2	1.3	1.5	1.5	0.6
lead-1	RR	1	55	0	-0.7	-17.2	1.2	1.7	3.4	4.1
	UR	0.8	46	-0.1	-0.6	-5.5	4.9	7.4	9.9	10.5
	ER	0.9	53	0	-0.7	-17.7	1.2	1.7	3.2	3.9
	Ad	0.9	51	0.1	19.2	-88.4	165.1	217.5	-2.5	-21.7
	TX	1.3	54	0.1	0.3	1.0	0.8	1	26.5	26.3
	TN	1	66	0.4	-0.1	-0.7	0.5	0.7	13.54	13.64
	TG	1.2	60	0.3	0.1	0.4	0.5	0.6	20	20
	GDD	1.2	58	0.2	1.3	0.6	11.1	13.8	210.3	20.9.0
	KDD	1.9	49	0.01	0.9	148.8	1.3	1.6	1.5	0.6
lead-2	RR	1	43	-0.2	-1	-23.9	1.4	1.9	3.1	4.1
	UR	0.9	51	-0.2	-0.7	-6.6	4.7	7.6	9.8	10.5
	ER	0.9	44	-0.2	-1	-24.6	1.4	1.9	3	3.9
	Ad	0.9	51	0.1	43.9	-202.0	166.9	222.7	22.2	-21.7
	TX	1.3	56	0.1	0.3	1.3	0.8	1	26.6	26.3
	TN	1	65	0.4	-0.1	-0.6	0.5	0.7	13.55	13.64
	TG	1.2	61	0.3	0.1	0.6	0.5	0.6	20.1	20
	GDD	1.3	58	0.3	2.6	1.3	10.7	13.4	211.7	209
	KDD	2.7	62	0.1	0.9	156.9	1.3	1.5	1.5	0.6

Table D.8: Forecast verification of Bungoma-Kenya reproductive stage climate indicators

	indicator	SprErr	GDS(%)	r	bias	pbias(%)	MAE	RMSE	meanfcst	meanobs
lead-0	UR	0.8	52	0	2.2	11.20	10	12.6	22.4	20.15
	ER	0.6	52	0.2	-0.2	-3.60	1.2	1.5	4.2	4.4
	Ad	0.8	54	0.1	-807.4	-168.70	1152.1	1482.2	-328.9	478.6
	TX	1.1	59	0.3	0.02	0.10	0.6	0.8	27.1	27.08
	TN	0.6	63	0.3	-0.13	-1.00	0.6	0.71	13.3	13.4
	TG	0.9	63	0.4	-0.1	-0.30	0.5	0.6	20.2	20.3
	GDD	1.4	51	-0.2	-3.6	-0.50	11.3	15.4	726.3	729.8
	KDD	1.3	61	0.2	1.4	12.20	7.6	10.1	13.3	11.9
lead-1	RR	0.6	55	0.1	-0.1	-1.70	1.2	1.5	4.4	4.4
	UR	0.9	48	0.1	1.7	8.40	9.7	11.8	21.8	20.2
	ER	0.6	55	0.1	-0.1	-1.50	1.2	1.5	4.2	4.4
	Ad	0.9	54	-0.03	-725.6	-151.60	1105.7	1458.4	-247.1	478.6
	TX	1.3	68	0.5	-0.02	-0.10	0.5	0.6	27.06	27.08
	TN	0.7	63	0.3	-0.1	-0.90	0.5	0.7	13.3	13.4
	TG	1	68	0.5	-0.1	-0.30	0.43	0.5	20.2	20.3
	GDD	1.5	48	-0.2	-5.1	-0.70	12.2	17.2	724.7	729.8
	KDD	1.4	60	0.2	1.2	10.30	7.2	8.9	13.1	11.9

Table D.9: Rank histogram test statistics for North-Gonder for lack of uniformity (X^2), and causes i.e. bias (Linear) and lack of variability (V-shape). In brackets is the p-value at 95% significance level. P-value<0.05 implies deviation from unity in case of X^2 or cause (in case of Linear and V-shape). The test-statistic in bolded where true.

		Growing season			vegetative stage			around anthesis			reproductive stage		
		χ^2	Linear	V-shape	χ^2	Linear	V-shape	χ^2	Linear	V-shape	χ^2	Linear	V-shape
lead-0	RR	11.6(0.71)	3.3(0.07)	0.04(0.84)	22.3(0.1)	0.6(0.45)	0.02(0.89)	7.3(0.94)	0.04(0.84)	0.03(0.9)	7.3(0.95)	0.001(0.1)	0.5(0.5)
	UR	20.1(0.17)	0.6(0.45)	1.7(0.2)	16.9(0.3)	0.3(0.6)	1.3(0.3)	11.6(0.71)	0.01(0.9)	2.2(0.14)	16.9(0.32)	0.2(0.7)	0.1(0.7)
	ER	10.5(0.78)	3.5(0.06)	0.09(0.76)	22.3(0.1)	0.3(0.6)	0.4(0.5)	12.6(0.6)	0.07(0.8)	0.05(0.8)	7.3(1.0)	0.2(0.7)	0.3(0.6)
	Ad	16.9(0.32)	1.9(0.17)	3.1(0.08)	19.1(0.2)	0.1(0.7)	2.8(0.09)	180(0.26)	0.23(0.6)	0.6(0.5)	7.3(1.0)	0.2(0.6)	0.01(0.9)
	TX	**37.2(0.00)**	**14.8(0.00)**	3.4(0.06)	**38.2(0.00)**	**9.1(0.00)**	3.3(0.07)	15.9(0.4)	6.0(0.01)	1.4(0.23)	**25.6(0.04)**	**7.0(0.00)**	1.1(0.3)
	TN	**31.8(0.01)**	2.5(0.11)	**13.0(0.00)**	**63.8(0.00)**	**8.6(0.00)**	14.2(0.00)	14.8(0.5)	0.7(0.4)	0.8(0.4)	19.1(0.2)	0.1(0.8)	1.1(0.3)
	TG	**38.3(8.2e-4)**	**17(3.8e-5)**	**5.4(0.02)**	**57.5(6.8e-7)**	**16.0(6.3e-5)**	**12.5(4.1e-4)**	21.2(0.1)	8.8(0.003)	0.1(0.8)	18.0(0.3)	6.2(0.01)	0.00(1.0)
	GDD	**90.0(0.00)**	**36.0(0.00)**	**16.1(0.00)**	**98.0(0.00)**	**30.7(0.00)**	**11.1(0.00)**	**25.0(0.03)**	**13.5(0.00)**	0.8(0.4)	20.1(0.2)	0.2(0.7)	1.0(0.3)
	KDD	**25.5(0.04)**	**7.6(0.01)**	**5.5(0.02)**	**25.5(0.04)**	2.4(0.12)	**2.1(0.15)**	11.6(0.7)	3.8(0.05)	2.2(0.14)	15.8(0.4)	4.4(0.04)	2.2(0.14)
lead-1	RR	12.7(0.63)	4.1(0.04)	0.1(0.77)	13.7(0.5)	2.0(0.14)	0.5(0.5)	10.5(0.8)	0.2(0.7)	1.2(0.3)	14.8(0.5)	3.5(0.06)	0.1(0.8)
	UR	10.5(0.78)	0.1(0.78)	0.4(0.54)	18.0(0.26)	0.9(0.34)	0.03(0.8)	20.1(0.2)	0.3(0.6)	0.83(0.4)	19.1(0.2)	0.5(0.5)	3.5(0.07)
	ER	14.8(0.47)	4.1(0.04)	0.003(0.9)	11.6(0.7)	1.9(0.2)	0.8(0.4)	10.5(0.8)	0.2(0.70	1.2(0.3)	16.9(0.3)	4.6(0.03)	0.0(1.0)
	Ad	12.7(0.63)	2.0(0.15)	0.43(0.5)	6.3(1.0)	0.01(0.9)	0.2(0.7)	14.8(0.5)	1.1(0.3)	1.0(0.3)	16.9(0.3)	2.9(0.1)	4.4(0.04)
	TX	19.1(0.21)	9.3(0.00)	0.4(0.5)	**26.5(0.03)**	**5.8(0.02)**	0.8(0.4)	15.9(0.4)	6.2(0.01)	1.1(0.3)	**35.1(0.002)**	**4.7(0.03)**	1.6(0.2)
	TN	**30.8(0.01)**	3.0(0.08)	**10.8(0.01)**	**85.2(0.00)**	**6.8(0.00)**	**17.5(0.00)**	19.1(0.2)	4.1(0.04)	1.3(0.3)	23.3(0.1)	1.9(0.2)	5.1(0.02)
	TG	23.3(0.08)	12.7(3.6e-4)	1.3(0.25)	**31.9(0.01)**	**9.1(0.002)**	**7.5(0.01)**	**25.5(0.04)**	7.7(0.01)	**9.7e-33(1)**	**24.4(0.06)**	**6.8(0.01)**	0.02(0.9)
	GDD	**38.0(0.00)**	**22.4(0.00)**	**3.7(0.05)**	**66.0(0.00)**	**20.7(0.00)**	**7.8(0.00)**	**29.7(0.01)**	**10.8(0.00)**	1.1(0.3)	**35.1(0.002)**	**25.3(4.9e-7)**	**5.4(0.02)**
	KDD	21.2(0.13)	3.8(0.05)	0.2(0.64)	16.9(0.32)	0.63(0.42)	0.3(0.56)	21.2(0.13)	3.0(0.08)	**0.13(0.7)**	14.8(0.5)	3.2(0.07)	0.1(0.8)
lead-2	RR	**27.6(0.02)**	**14.8(0.00)**	**4.3(0.04)**	19.1(0.21)	10.0(0.00)	3.7(0.1)	18.0(0.3)	10.3(0.00)	0.1(0.8)	14.8(0.5)	0.2(0.7)	0.6(0.4)
	UR	14.8(0.5)	0.01(0.93)	1.3(0.25)	15.9(0.4)	2.0(0.15)	0.6(0.5)	14.8(0.5)	0.01(0.9)	2.5(0.1)	9.5(0.9)	3.6(0.06)	0(1.0)
	ER	**26.5(0.03)**	**12.7(0.00)**	2.7(0.1)	**31.9(0.01)**	**8.4(0.00)**	2.8(0.1)	16.9(0.3)	9.3(0.00)	0.1(0.8)	7.3(1.0)	0.03(0.90	0.4(0.6)
	Ad	10.5(0.78)	1.3(0.25)	1.3(0.3)	18.0(0.3)	5.8(0.02)	1.1(0.3)	21.2(0.13)	0.03(0.87)	2.8(0.09)	15.9(0.4)	3.6(0.06)	0.01(0.9)
	TX	**27.6(0.02)**	**14.8(0.00)**	1.8(0.2)	21.2(0.13)	9.3(0.00)	1.0(0.3)	**26.5(0.03)**	15.7(0.00)	1.1(0.3)	22.3(0.1)	8.6(0.003)	0.3(0.6)
	TN	**31.9(0.01)**	1.5(0.2)	**12.1(0.00)**	**64.9(0.00)**	**7.3(0.01)**	**18.2(0.00)**	18.0(0.3)	0.7(0.4)	1.0(0.3)	10.5(0.8)	0.001(1.0)	0.1(0.8)
	TG	24.4(0.06)	**13.9(2.0e-4)**	0.6(0.4)	**27.6(0.02)**	**15.4(8.8e-5)**	3.0(0.08)	23.3(0.08)	14.5(1.4e-4)	0.8(0.4)	**27.6(0.02)**	**7.3(0.01)**	0.4(0.5)
	GDD	**391.0(0.00)**	**78.0(0.00)**	**97.5(0.00)**	19.1(0.2)	13.6(0.00)	0.001(0.96)	**34.0(0.00)**	**15.4(0.00)**	1.9(0.2)	**30.8(0.01)**	**12.7(3.6e-4)**	2.6(0.1)
	KDD	18.0(0.26)	3.6(0.06)	0.9(0.33)	11.6(0.7)	3.2(0.07)	0.48(0.5)	22.2(0.1)	9.3(0.00)	0.4(0.6)	22.3(0.1)	5.5(0.02)	3.0(0.09)

Table D.10: Same as supplementary Table D.9 but for Bungoma

		Growing season			vegetative stage			around anthesis			reproductive stage		
		χ^2	Linear	V-shape	χ^2	Linear	V-shape	χ^2	Linear	V-shape	χ^2	Linear	V-shape
lead-0	RR	18.0(0.3)	0.8(0.4)	1.8(0.2)	16.9(0.3)	1.9(0.2)	2.5(0.1)	**24.4(0.05)**	**14.8(0.00)**	0.5(0.5)	**27.6(0.02)**	**0.01(0.9)**	**13.3(2.6e-4)**
	UR	19.1(0.2)	0.8(0.4)	**11.4(7.2e-4)**	**45.7(5.9e-5)**	**3.6(0.06)**	1.6(0.2)	12.7(0.6)	0.2(0.7)	1.0(0.3)	**28.7(0.02)**	2.4(0.1)	**5.4(0.02)**
	ER	22.3(0.1)	0.8(0.4)	3.3(0.07)	14.8(0.5)	1.4(0.2)	2.5(0.1)	20.1(0.2)	12.7(3.6e-4)	0.4(0.5)	**31.9(0.01)**	0.04(0.8)	**13.3(2.6e-4)**
	Ad	**26.5(0.03)**	0.01(0.9)	**11.8(5.8e-4)**	19.1(0.2)	4.2(0.04)	0.8(0.4)	10.5(0.8)	0.04(0.8)	0.04(0.8)	20.1(0.2)	12.7(3.6e-4)	1.0(0.3)
	TX	11.6(0.7)	0.8(0.4)	0.01(0.9)	12.7(0.6)	0.2(0.7)	7.7e-34(1)	20.1(0.2)	3.3(0.07)	0.4(0.5)	13.7(0.5)	0.03(0.9)	0.1(0.8)
	TN	**37.2(0.001)**	**4.2(0.04)**	**15.8(6.9e-5)**	**27.6(0.02)**	**6.8(0.01)**	**8.4(0.003)**	13.7(0.5)	1.1(0.7)	0.2(0.7)	14.8(0.5)	2.3(0.1)	9.7(0.00)
	TG	10.5(0.8)	1.3(0.3)	1.9(0.2)	21.2(0.13)	2.8(0.1)	3.8(0.05)	13.7(0.5)	1.9(0.2)	1.0(0.3)	16.9(0.3)	1.1(0.3)	2.6(0.1)
	GDD	**32.9(0.004)**	**11.1(8.7e-4)**	2.6(0.1)	15.9(0.4)	5.3(0.02)	1.6(0.21)	13.7(0.5)	1.6(0.2)	1.0(0.3)	20.1(0.2)	1.9(0.2)	8.7(0.0030
	KDD	6.3(0.97)	0.2(0.6)	0.04(0.8)	16.9(0.3)	0.8(0.4)	0.0(0.9)	10.5(0.8)	1.0(0.3)	2.8(0.1)	11.6(0.7)	0.13(0.7)	0.05(0.8)
lead-1	RR	9.5(0.9)	1e-32(0.00)	1.3(0.3)	18.0(0.3)	0.01(0.9)	0.2(0.6)	20.1(0.17)	6.6(0.01)	0.15(0.7)	14.8(0.5)	0.8(0.4)	6.5(0.01)
	UR	19.1(0.2)	1.0(0.3)	5.0(0.03)	20.1(0.2)	2.8(0.1)	3.2(0.07)	12.6(0.6)	0.03(0.8)	0.1(0.8)	**27.6(0.02)**	1.3(0.3)	**4.4(0.04)**
	ER	10.5(0.8)	4.0e-34(1)	3.2(0.07)	10.5(0.8)	0.08(0.8)	0.0(1.0)	19.1(0.2)	5.6(0.02)	0.1(0.7)	15.9(0.4)	0.6(0.5)	7.5(0.01)
	Ad	12.7(0.6)	0.5(0.5)	7.7(0.01)	22.3(0.1)	2.0(0.2)	0.3(0.6)	13.7(0.5)	0.03(0.9)	0.3(0.6)	**26.5(0.03)**	**14.2(1.7e-4)**	0.3(0.6)
	TX	10.5(0.8)	1.0(0.3)	2.2(0.1)	12.7(0.6)	0.03(0.9)	4.9(0.03)	15.9(0.4)	1.0(0.3)	1.0(0.3)	6.5(0.9)	0.2(0.6)	2.4(0.1)
	TN	**42.5(1.9e-4)**	3.5(0.06)	**15.1(1e-4)**	**25.5(0.04)**	3.0(0.08)	**11.4(7.2e-4)**	10.5(0.8)	0.98(0.3)	0.4(0.5)	20.1(0.2)	3.2(0.1)	5.4(0.02)
	TG	23.3(0.08)	2.6(0.1)	1.5(0.2)	9.5(0.9)	1.6(0.2)	0.4(0.5)	18.0(0.3)	0.1(0.8)	0.1(0.7)	9.5(0.9)	0.6(0.4)	0.01(1.0)
	GDD	**38.3(8.2e-4)**	**13.9(2.0e-4)**	**3.8(0.05)**	**25.5(0.04)**	**4.2(0.04)**	1.5(0.2)	11.6(0.7)	0.01(0.9)	0.00(1.0)	18.0(0.3)	3.9(0.05)	7.9(0.01)
	KDD	19.1(0.2)	0.1(0.8)	0.2(0.7)	10.5(0.8)	1.0(0.3)	4.0(0.04)	20.1(0.2)	2.8(0.1)	0.3(0.6)	21.2(0.13)	0.01(0.9)	0.12(0.7)
lead-2	RR				9.5(0.9)	0.19(0.7)	3e-34(1)	15.9(0.4)	8.8(0.003)	0.1(0.7)			
	UR				20.1(0.2)	2.3(0.1)	0.6(0.5)	9.5(0.9)	0.1(0.7)	1.2(0.3)			
	ER				10.5(0.8)	0.2(0.7)	0.1(0.7)	15.9(0.4)	9.8(0.002)	0.1(0.7)			
	Ad				12.7(0.6)	2.0(0.2)	0.3(0.6)	18.0(0.3)	0.3(0.6)	0.1(0.8)			
	TX				14.8(0.5)	0.002(1.0)	1.9(0.2)	14.8(0.5)	1.5(0.2)	1.8(0.2)			
	TN				**25.5(0.04)**	**4.1(0.04)**	**11.0(8.9e-4)**	14.8(0.5)	0.2(0.7)	0.5(0.5)			
	TG				14.8(0.5)	2.1(0.1)	0.7(0.4)	11.6(0.7)	0.3(0.6)	1.5(0.2)			
	GDD				30.8(0.01)	3.3(0.07)	3.2(0.08)	13.7(0.5)	0.4(0.5)	0.9(0.3)			
	KDD				12.7(0.6)	1.9(0.2)	0.1(0.70	1.7(0.5)	0.8(0.4)	3.5(0.1)			

D.2 Chapter-6 supplementary tables

Table D.11: Bungoma NUT2 Yield and TAGP predictability aggregated by dominant maize variety and soil type at FAO soil mapping units (0.1° resolution).Predictability is shown using ROCSS averaged over grid cells with significant skill scores. In brackets are percentage of grid cells with significant ROCSS. Soil type S3 mean “LOAMY SAND-to-SANDY LOAM” , and S4 means “SANDY LOAM-to-LOAM soils.

BUNGOMA VARIETY-6; S3-SOILS

	YIELDS			TAGP		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
BN	0	0.1	-0.2	0.2	0.2	0
NN	0.3	-0.2	0	0.1	-0.1	0
AN	0	0.2	0.2	-0.1	-0.1	-0.1

Table D.12: North Gonder NUT2 Yield predictability aggregated by dominant maize variety and soil type at FAO soil mapping units (0.1°resolution).Predictability is shown using ROCSS averaged over grid cells with significant skill scores. In brackets are percentage of grid cells with significant ROCSS. Soil type S3 means "LOAMY SAND-to-SANDY LOAM" , and S4 means "SANDY LOAM-to-LOAM soils

NORTH GONDER (0.1°RESOLUTION)
AGGREGATED YIELDS TO NUT2 REGION BY SOIL TYPE

(a)

	VARIETY-5								
	S2 SOILS			S3 SOILS			S4 SOILS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
AN	0.1	0	-	0.1	0	-	0.1	0	-
NN	0	-0.5	-	-0.2	-0.5	-	-0.2	-0.5	-
BN	0.5* (100%)	0.4* (100%)	-	0.5* (100%)	0.5* (100%)	-	0.5* (100%)	0.6* (100%)	-

(b)

	VARIETY-6								
	S2 SOILS			S3 SOILS			S4 SOILS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
AN	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
NN	-0.3	-0.1	-0.3	-0.3	-0.1	-0.2	-0.3	-0.1	-0.3
BN	0.4* (100%)	0.5* (100%)	0.3	0.4* (100%)	0.5* (100%)	0.3	0.3* (100%)	0.5* (100%)	0.3

(c)

	VARIETY-7								
	S2 SOILS			S3 SOILS			S4 SOILS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
AN	0.5* (100%)	0.3	-	0.5* (100%)	0.3	-	0.5* (100%)	0.3	-
NN	0.3	-0.1	-	0.4* (81%)	-0.1	-	0.4* (100%)	0	-
BN	0.7* (100%)	0.6* (100%)	-	0.7* (100%)	0.6* (100%)	-	0.8* (100%)	0.7* (100%)	-

(d)

	VARIETY-10								
	S2 SOILS			S3 SOILS			S4 SOILS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
AN	0.1	0	-0.1	-0.1	-0.2	-0.1	0.3	0.3	0.3
NN	-0.2	-0.1	-0.1	0	-0.3	0.1	0.1	0.3	0.3
BN	0.3	0.3	0.3	0.1	0	0.2	0.1	0.1	0.2

* shows significant skill at 95% level

Table D.13: North Gonder NUT2 TAGP predictability aggregated by dominant maize variety and soil type at FAO soil mapping units (0.1°resolution).Predictability is shown using ROCSS averaged over grid cells with significant skill scores. In brackets are percentage of grid cells with significant ROCSS. Soil type S3 means "LOAMY SAND-to-SANDY LOAM" , and S4 means "SANDY LOAM-to-LOAM soils

NORTH GONDER (0.1° RESOLUTION)
AGGREGATED TAGP TO NUT2 REGION BY SOIL TYPE

(a) VARIETY-5

	S2 SOILS			S3 SOILS			S4 SOILS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
AN	0.3	0.3	-	0.3	0.3	-	0.3	0.3	-
NN	-0.1	0.1	-	0	0.1	-	0	0.2	-
BN	0.35	0.3	-	0.4* (100%)	0.3	-	0.4* (100%)	0.3	-

(b) VARIETY-6

	S2 SOILS			S3 SOILS			S4 SOILS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
AN	0.1	0.3	0.1	-0.2	0.3	0.1	0.2	0.3	0.1
NN	-0.2	0.2	-0.2	-0.1	0.2	-0.2	-0.2	0.2	-0.2
BN	0	0.2	0.1	0	0.2	0.2	0	0.2	0.2

(c) VARIETY-7

	S2 SOILS			S3 SOILS			S4 SOILS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
AN	0.4* (100%)	0.3	-	0.4* (100%)	0.3	-	0.4* (100%)	0.3	-
NN	-0.2	0.1	-	-0.2	0.2	-	-0.2	0.2	-
BN	0.3	0.4* (100%)	-	0.3	0.4* (100%)	-	0.3	0.4	-

(d) VARIETY-10

	S2 SOILS			S3 SOILS			S4 SOILS		
	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2	LEAD-0	LEAD-1	LEAD-2
AN	0.3	0.2	0.2	0.3 (45%)	0.2	0.3	0.5* (100%)	0.6* (100%)	0.6* (100%)
NN	0	0.3	-0.2	0.1	0.3* (72%)	0.3* (18%)	0.1	0.3	0.1
BN	.0.2	0.2	0	0.2	0.2	0.1	0.7* (100%)	0.6* (100%)	0.5* (100%)

* shows significant skill at 95% level

References

Adimo, O, A., 2020: Description of cropping systems, climate and soils in kenya. Avaialable at `http://www.yieldgap.org/kenya`(2020/05/06).

AGRA, 2019: Feeding africa's soils: Fertilizers to support africa's agricultural transformations. Technical report, Nairobi,Kenya.

Amengual, A., V. Homar, R. Romero, S. Alonso, and C. Ramis, 2012: Projections of the climate potential for tourism at local scales: application to platja de palma, spain. *International Journal of Climatology*, **32**(14), 2095–2107.

Amissah-Arthur, A., S. Jagtap, and C. Rosenzweig, 2002: Spatio-temporal effects of el niño events on rainfall and maize yield in kenya. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **22**(15), 1849–1860.

Anderson, C. J., B. A. Babcock, Y. Peng, P. W. Gassman, and T. D. Campbell, 2015: Placing bounds on extreme temperature response of maize. *Environmental Research Letters*, **10**(12), 124001.

Armstrong, C., 1999: The first decade of maize transformation: a review and future perspective. *Maydica*, **44**(1), 101–109.

Asseng, S., F. Ewert, C. Rosenzweig, J. W. Jones, J. L. Hatfield, A. C. Ruane, K. J. Boote, P. J. Thorburn, R. P. Rötter, D. Cammarano, et al., 2013: Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, **3**(9), 827–832.

Asseng, S., N. Turner, and B. A. Keating, 2001: Analysis of water-and nitrogen-use efficiency of wheat in a mediterranean climate. *Plant and Soil*, **233**(1), 127–143.

Auffhammer, M., 2011: Weather dilemma for african maize. *Nature Climate Change*, **1**(1), 27–28.

Bahaga, T., F. Kucharski, G. M. Tsidu, and H. Yang, 2016: Assessment of prediction and predictability of short rains over equatorial east africa using a multi-model ensemble. *Theoretical and Applied Climatology*, **123**(3-4), 637–649.

Baigorria, G. A., J. W. Jones, and J. J. O'Brien, 2008: Potential predictability of crop yield using an ensemble climate forecast by a regional circulation model. *Agricultural and Forest Meteorology*, **148**(8-9), 1353–1361.

Baigorria, G. A., J. W. Jones, D.-W. Shin, A. Mishra, and J. J. O'Brien, 2007: Assessing uncertainties in crop model simulations using daily bias-corrected regional circulation model outputs. *Climate Research*, **34**(3), 211–222.

Baldocchi, D., E. Falge, L. Gu, R. Olson, D. Hollinger, S. Running, P. Anthoni, C. Bernhofer, K. Davis, R. Evans, et al., 2001: Fluxnet: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, **82**(11), 2415–2434.

Bannayan, M., S. Sanjani, A. Alizadeh, S. S. Lotfabadi, and A. Mohamadian, 2010: Association between climate indices, aridity index, and rainfed crop yield in northeast of iran. *Field Crops Research*, **118**(2), 105–114.

Barnston, A. G., S. Li, S. J. Mason, D. G. DeWitt, L. Goddard, and X. Gong, 2010: Verification of the first 11 years of iri's seasonal climate forecasts. *Journal of Applied Meteorology and Climatology*, **49**(3), 493–520.

Baron, C., B. Sultan, M. Balme, B. Sarr, S. Traore, T. Lebel, S. Janicot, and M. Dingkuhn, 2005: From gcm grid cell to agricultural plot: scale issues affecting modelling of climate impact. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**(1463), 2095–2108.

Bassu, S., N. Brisson, J.-L. Durand, K. Boote, J. Lizaso, J. W. Jones, C. Rosenzweig, A. C. Ruane, M. Adam, C. Baron, et al., 2014: How do various maize crop models vary in their responses to climate change factors? *Global Change Biology*, **20**(7), 2301–2320.

Batjes, N. H., 2012: Isric-wise derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.2). Technical report, ISRIC-World Soil Information.

Bechtold, P., M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. J. Rodwell, F. Vitart, and G. Balsamo, 2008: Advances in simulating atmospheric variability with the ecmwf model: From synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, **134**(634), 1337–1351.

Bhend, J., J. Ripoldi, C. Mignani, I. Mahlstein, R. Hiller, C. Spirig, M. Liniger, A. Weigel, J. Bedia Jimenez, M. De Felice, et al., 2016: easyverification: Ensemble forecast verification for large data sets.

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

Black, E., J. Slingo, and K. R. Sperber, 2003: An observational study of the relationship between excessively strong short rains in coastal east africa and indian ocean sst. *Monthly Weather Review*, **131**(1), 74–94.

Bonilla Cedrez, C., J. Chamberlin, Z. Guo, and R. J. Hijmans, 2020: Spatial variation in fertilizer prices in sub-saharan africa. *PloS one*, **15**(1), e0227764.

Boogaard, H., J. Wolf, I. Supit, S. Niemeyer, and M. van Ittersum, 2013: A regional implementation of wofost for calculating yield gaps of autumn-sown wheat across the european union. *Field Crops Research*, **143**, 130–142.

Borhara, K., B. Pokharel, B. Bean, L. Deng, and S.-Y. S. Wang, 2020: On tanzania's precipitation climatology, variability, and future projection. *Climate*, **8**(2), 34.

Boucher, O. and M. Best, 2010: The watch forcing data 1958-2001: A meteorological forcing dataset for land surface-and hydrological-models. Technical report.

Boulanger, P., H. Dudu, E. Ferrari, A. M. Causapé, J. Balié, and L. Battaglia, 2018: Policy options to support the agriculture sector growth and transformation strategy in kenya. *A CGE Analysis, EUR*, **29231**.

Brown, M. E., C. C. Funk, G. Galu, and R. Choularton, 2007: Earlier famine warning possible using remote sensing and models. *Eos, Transactions American Geophysical Union*, **88**(39), 381–382.

Buizza, R. and T. N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Monthly Weather Review*, **126**(9), 2503–2518.

Butler, E. E. and P. Huybers, 2013: Adaptation of us maize to temperature variations. *Nature Climate Change*, **3**(1), 68–72.

Butler, E. E. and P. Huybers, 2015: Variations in the sensitivity of us maize yield to extreme temperatures by region and growth phase. *Environmental Research Letters*, **10**(3), 034009.

Butts-Wilmsmeyer, C. J., J. R. Seebauer, L. Singleton, and F. E. Below, 2019: Weather during key growth stages explains grain quality and yield of maize. *Agronomy*, **9**(1), 16.

C3S, C. C. C. S., 2017: Era5: Fifth generation of ecmwf atmospheric reanalyses of the global climate.

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

Cakir, R., 2004: Effect of water stress at different development stages on vegetative and reproductive growth of corn. *Field Crops Research*, **89**(1), 1–16.

Camberlin, P., S. Janicot, and I. Poccard, 2001: Seasonality and atmospheric dynamics of the teleconnection between african rainfall and tropical sea-surface temperature: Atlantic vs. enso. *International Journal of Climatology*, **21**(8), 973–1005.

Camberlin, P. and N. Philippon, 2002: The east african march–may rainy season: Associated atmospheric dynamics and predictability over the 1968–97 period. *Journal of Climate*, **15**(9), 1002–1019.

Cane, M. A., G. Eshel, and R. W. Buckland, 1994: Forecasting zimbabwean maize yield using eastern equatorial pacific sea surface temperature. *Nature*, **370**(6486), 204–205.

Cantelaube, P. and J.-M. Terres, 2005: Seasonal weather forecasts for crop yield modelling in europe. *Tellus A: Dynamic Meteorology and Oceanography*, **57**(3), 476–487.

Ceglar, A. and L. Kajfež-Bogataj, 2012: Simulation of maize yield in current and changed climatic conditions: addressing modelling uncertainties and the importance of bias correction in climate model simulations. *European Journal of Agronomy*, **37**(1), 83–95.

Challinor, A., J. Slingo, T. Wheeler, and F. Doblas-Reyes, 2005: Probabilistic simulations of crop yield over western india using the demeter seasonal hindcast ensembles. *Tellus A: Dynamic Meteorology and Oceanography*, **57**(3), 498–512.

Charles, A., Y. Kuleshov, and D. Jones, 2012: Managing climate risk with seasonal forecasts. In *Risk Management-Current Issues and Challenges*. IntechOpen.

Charles, A. N., R. E. Duell, X. Wang, and A. B. Watkins, 2015: Seasonal forecasting for australia using a dynamical model: improvements in forecast skill over the operational statistical model. *Australian Meteorological and Oceanographic Journal*, **65**(3), 356–375.

Cioffi, F., F. Conticello, and U. Lall, 2016: Projecting changes in tanzania rainfall for the 21st century. *International Journal of Climatology*, **36**(13), 4297–4314.

Cucchi, M., G. P. Weedon, A. Amici, N. Bellouin, S. Lange, H. M. Schmied, H. Hersbach, and C. Buontempo, 2020: Wfde5: bias adjusted era5 reanalysis data for impact studies. *Earth System Science Data Discussions*, **2020**, 1–32.

De Ridder, N., M. Van Ittersum, and H. Van Keulen, 2000: Four explorative land use studies in south and southeast asia at the ecoregional scale: How future-oriented are they? *Systems research for optimizing future land use in South and Southeast Asia* 165.

de Wit, A., B. Baruth, H. Boogaard, K. van Diepen, D. van Kraalingen, F. Micale, J. te Roller, I. Supit, and R. van den Wijngaart, 2010: Using era-interim for regional crop yield forecasting in europe. *Climate Research*, **44**(1), 41–53.

de Wit, A., H. Boogaard, D. Fumagalli, S. Janssen, R. Knapen, D. van Kraalingen, I. Supit, R. van der Wijngaart, and K. van Diepen, 2019: 25 years of the wofost cropping systems model. *Agricultural Systems*, **168**, 154–167.

Declaration, A., 2006: Abuja declaration on fertilizer for the african green revolution, issued at the africa fertilizer summit: African union special summit of the heads of state and government. *Abuja, Nigeria*, **13**.

Dee, D. P., S. M. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, d. P. Bauer, et al., 2011: The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**(656), 553–597.

Deryng, D., W. Sacks, C. Barford, and N. Ramankutty, 2011: Simulating the effects of climate and agricultural management practices on global crop yield. *Global Biogeochemical Cycles*, **25**(2).

Dinku, T., P. Ceccato, E. Grover-Kopec, M. Lemma, S. Connor, and C. Ropelewski, 2007: Validation of satellite rainfall products over east africa's complex topography. *International Journal of Remote Sensing*, **28**(7), 1503–1526.

Diro, G., A. Tompkins, and X. Bi, 2012: Dynamical downscaling of ecmwf ensemble seasonal forecasts over east africa with regcm3. *Journal of Geophysical Research: Atmospheres*, **117**(D16).

Diro, G. T., D. Grimes, and E. Black, 2011: Large scale features affecting ethiopian rainfall. In *African climate and climate change*. Springer, 13–50.

Doblas-Reyes, F., V. Pavan, and D. Stephenson, 2003: The skill of multi-model seasonal forecasts of the wintertime north atlantic oscillation. *Climate Dynamics*, **21**(5-6), 501–514.

Doblas-Reyes, F., A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J. Murphy, P. Rogel, D. Smith, and T. Palmer, 2009: Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, **135**(643), 1538–1559.

Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. Rodrigues, 2013: Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, **4**(4), 245–268.

Duffy, K. J. and T. P. Masere, 2015: Effect of within-season daily rainfall distribution on maize crop yields. *Outlook on Agriculture*, **44**(4), 267–271.

Dutra, E., L. Magnusson, F. Wetterhall, H. L. Cloke, G. Balsamo, S. Boussetta, and F. Pappenberger, 2013: The 2010–2011 drought in the horn of africa in ecmwf reanalysis and seasonal forecast products. *International Journal of Climatology*, **33**(7), 1720–1729.

Dutra, E., W. Pozzi, F. Wetterhall, F. Di Giuseppe, L. Magnusson, G. Naumann, P. Barbosa, J. Vogt, and F. Pappenberger, 2014: Global meteorological drought-part 2: Seasonal forecasts. *Hydrology & Earth System Sciences*, **18**(7).

Ebert, E., L. Wilson, A. Weigel, M. Mittermaier, P. Nurmi, P. Gill, M. Göber, S. Joslyn, B. Brown, T. Fowler, et al., 2013: Progress and challenges in forecast verification. *Meteorological Applications*, **20**(2), 130–139.

Elmore, K. L., 2005: Alternatives to the chi-square test for evaluating rank histograms from ensemble forecasts. *Weather and Forecasting*, **20**(5), 789–795.

Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, **8**(6), 985–987.

Fantaye, K., 2020: Description of cropping systems, climate and soils in ethiopia. Available at `http://www.yieldgap.org/ethiopia`[2020/05/06].

FAO Global information and early warning system. Technical report, FAO, Rome.

FAO, 2020a: Fao in ethiopia.

FAO, 2020b: Fao in kenya.

Fischer, G., F. Nachtergaele, S. Prieler, H. Van Velthuizen, L. Verelst, and D. Wiberg, 2008: Global agro-ecological zones assessment for agriculture (gaez 2008). *IIASA, Laxenburg, Austria and FAO, Rome, Italy*, **10**.

Fisher, A. C., W. M. Hanemann, M. J. Roberts, and W. Schlenker, 2012: The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: comment. *American Economic Review*, **102**(7), 3749–60.

Folberth, C., T. Gaiser, K. C. Abbaspour, R. Schulin, and H. Yang, 2012: Regionalization of a large-scale crop growth model for sub-saharan africa: Model setup, evaluation, and estimation of maize yields. *Agriculture, Ecosystems & Environment*, **151**, 21–33.

Funk, C., S. E. Nicholson, M. Landsfeld, D. Klotter, P. Peterson, and L. Harrison, 2015: The centennial trends greater horn of africa precipitation dataset. *Scientific Data*, **2**(1), 1–17.

Funk, C., P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, et al., 2015: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, **2**(1), 1–21.

Funk, C., P. Peterson, S. Peterson, S. Shukla, F. Davenport, J. Michaelsen, K. R. Knapp, M. Landsfeld, G. Husak, L. Harrison, et al., 2019: A high-resolution 1983–2016 t max climate data record based on infrared temperatures and stations by the climate hazard center. *Journal of Climate*, **32**(17), 5639–5658.

Gitau, W., P. Camberlin, L. Ogallo, and R. Okoola, 2015: Oceanic and atmospheric linkages with short rainfall season intraseasonal statistics over equatorial eastern africa and their predictive potential. *International Journal of Climatology*, **35**(9), 2382–2399.

Glotter, M., J. Elliott, D. McInerney, N. Best, I. Foster, and E. J. Moyer, 2014: Evaluating the utility of dynamical downscaling in agricultural impacts projections. *Proceedings of the National Academy of Sciences*, **111**(24), 8776–8781.

Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology*, **21**(9), 1111–1152.

GOK, 2019: Agriculture sector transformation and growth strategy. Technical report, Ministry of Agriculture, Livestock, Fisheries and Irrigation.

Graham, R., A. Evans, K. Mylne, M. Harrison, and K. Robertson, 2000: An assessment of seasonal predictability using atmospheric general circulation models. *Quarterly Journal of the Royal Meteorological Society*, **126**(567), 2211–2240.

Guo, R., Z. Lin, X. Mo, and C. Yang, 2010: Responses of crop yield and water use efficiency to climate change in the north china plain. *Agricultural Water Management*, **97**(8), 1185–1194.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, **129**(3), 550–560.

Hamill, T. M. and J. Juras, 2006: Measuring forecast skill: is it real skill or is it the varying climatology? *Quarterly Journal of the Royal Meteorological Society*, **132**(621C), 2905–2923.

Hansen, J. W., A. Challinor, A. Ines, T. Wheeler, and V. Moron, 2006: Translating climate forecasts into agricultural terms: advances and challenges. *Climate Research*, **33**(1), 27–41.

Hansen, J. W. and M. Indeje, 2004: Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid kenya. *Agricultural and Forest Meteorology*, **125**(1-2), 143–157.

Hansen, J. W. and A. V. Ines, 2005: Stochastic disaggregation of monthly rainfall data for crop simulation studies. *Agricultural and Forest Meteorology*, **131**(3-4), 233–246.

Hansen, J. W., S. J. Mason, L. Sun, and A. Tall, 2011: Review of seasonal climate forecasting for agriculture in sub-saharan africa. *Experimental Agriculture*, **47**(2), 205–240.

Hansen, J. W., A. Mishra, K. Rao, M. Indeje, and R. K. Ngugi, 2009: Potential value of gcm-based seasonal rainfall forecasts for maize management in semi-arid kenya. *Agricultural Systems*, **101**(1-2), 80–90.

Harding, R., M. Best, E. Blyth, S. Hagemann, P. Kabat, L. M. Tallaksen, T. Warnaars, D. Wiberg, G. P. Weedon, H. v. Lanen, et al., 2011: Watch: Current knowledge of the terrestrial global water cycle. *Journal of Hydrometeorology*, **12**(6), 1149–1156.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2014: Updated high-resolution grids of monthly climatic observations–the cru ts3. 10 dataset. *International Journal of Climatology*, **34**(3), 623–642.

Harrison, M., 2005: The development of seasonal and inter-annual climate forecasting. In *Increasing Climate Variability and Change*. Springer, 201–220.

Hastenrath, S., 1995: Recent advances in tropical climate prediction. *Journal of Climate*, **8**(6), 1519–1532.

Hatfield, J. L., K. J. Boote, B. Kimball, L. Ziska, R. C. Izaurralde, D. Ort, A. M. Thomson, and D. Wolfe, 2011: Climate impacts on agriculture: implications for crop production. *Agronomy Journal*, **103**(2), 351–370.

Hatfield, J. L. and C. Dold, 2018: Climate change impacts on corn phenology and productivity. *Corn: Production and Human Health in Changing Climate* 95.

Hatfield, J. L. and J. H. Prueger, 2015: Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, **10**, 4–10.

Hengsdijk, H., G. Meijerink, and M. Mosugu, 2005: Modeling the effect of three soil and water conservation practices in tigray, ethiopia. *Agriculture, Ecosystems & Environment*, **105**(1-2), 29–40.

Hersbach, H., 2016: The era5 atmospheric reanalysis. In *AGU fall meeting abstracts*.

Hillocks, R., 2014: Addressing the yield gap in sub-saharan africa. *Outlook on Agriculture*, **43**(2), 85–90.

Hoell, A. and C. Funk, 2014: Indo-pacific sea surface temperature influences on failed consecutive rainy seasons over eastern africa. *Climate Dynamics*, **43**(5-6), 1645–1660.

Hoffmann, H., K. C. Kersebaum, G. Zhao, S. Asseng, M. Bindi, D. Cammarano, J. Constantin, E. Coucheney, R. Dechow, L. Doro, et al., 2016: Data aggregation effects in regional yield simulations.

Hoskins, B., 2006: *Predictability from a dynamical meteorological perspective*. Cambridge University Press, 30–39.

Huang, C., S. W. Duiker, L. Deng, C. Fang, and W. Zeng, 2015: Influence of precipitation on maize yield in the eastern united states. *Sustainability*, **7**(5), 5996–6010.

Huang, J., X. Chen, L. Zhou, Y. Xue, and J. Lin, 2017: Statistical analysis of the relationship between climate-induced maize yield and rainy-season precipitation across inner mongolia, north china. *Theoretical and Applied Climatology*, **129**(3-4), 1145–1156.

Hui, J., E.-d. LIN, T. Wheeler, A. Challinor, and S. Jiang, 2013: Climate change modelling and its roles to chinese crops yield. *Journal of Integrative Agriculture*, **12**(5), 892–902.

Husak, G. J., J. Michaelsen, P. Kyriakidis, J. P. Verdin, C. Funk, and G. Galu, 2011: The forecast interpretation tool—a monte carlo technique for blending climatic distributions with probabilistic forecasts. *International Journal of Climatology*, **31**(3), 461–467.

Huzsvay, L. and J. Nagy, 2005: Effect of weather on maize yields and the efficiency of fertilization. *Acta Agronomica Hungarica*, **53**(1), 31–39.

ICPAC, 2018: . The fiftieth Greater Horn of Africa climate outlook forum (GHA-COF50) bulletin October-December 2028.

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

Iizumi, T., H. Sakuma, M. Yokozawa, J.-J. Luo, A. J. Challinor, M. E. Brown, G. Sakurai, and T. Yamagata, 2013: Prediction of seasonal climate-induced variations in global food production. *Nature Climate Change*, **3**(10), 904–908.

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

IPCC, 2014: Climate change 2014: impacts, adaptation, and vulnerability. part a: global and sectoral aspects. contribution of working group ii to the fifth assessment report of the intergovernmental panel on climate change. 1132.

Jakob Themeßl, M., A. Gobiet, and A. Leuprecht, 2011: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, **31**(10), 1530–1544.

Jan van Oldenborgh, G., M. A. Balmaseda, L. Ferranti, T. N. Stockdale, and D. L. Anderson, 2005: Did the ecmwf seasonal forecast model outperform statistical enso forecast models over the last 15 years? *Journal of Climate*, **18**(16), 3240–3249.

Johnson, S. J., T. N. Stockdale, L. Ferranti, M. A. Balmaseda, F. Molteni, L. Magnusson, S. Tietsche, D. Decremer, A. Weisheimer, G. Balsamo, et al., 2019: Seas5: the new ecmwf seasonal forecast system. *Geoscientific Model Development*, **12**(3).

Jolliffe, I. T. and C. Primo, 2008: Evaluating rank histograms using decompositions of the chi-square test statistic. *Monthly Weather Review*, **136**(6), 2133–2139.

Jones, J. W., J. W. Hansen, F. S. Royce, and C. D. Messina, 2000: Potential benefits of climate forecasting to agriculture. *Agriculture, Ecosystems & Environment*, **82**(1-3), 169–184.

Jung, T., G. Balsamo, P. Bechtold, A. Beljaars, M. Koehler, M. Miller, J.-J. Morcrette, A. Orr, M. Rodwell, and A. M. Tompkins, 2010: The ecmwf model climate: recent progress through improved physical parametrizations. *Quarterly Journal of the Royal Meteorological Society*, **136**(650), 1145–1160.

Jury, M. R. and C. Funk, 2013: Climatic trends over ethiopia: regional signals and drivers. *International Journal of Climatology*, **33**(8), 1924–1935.

Kadi, H. A. K., L. N. Njau, J. Mwikya, and A. Kamga, 2011: The state of climate information services for agriculture and food security in east african countries.

Kalnay, E., B. Hunt, E. Ott, and I. Szunyogh, 2006: *Ensemble forecasting and data assimilation: two problems with the same solution?* Cambridge University Press, 157–180.

Kamali, B., K. C. Abbaspour, A. Lehmann, B. Wehrli, and H. Yang, 2018: Spatial assessment of maize physical drought vulnerability in sub-saharan africa: Linking drought exposure with crop failure. *Environmental Research Letters*, **13**(7), 074010.

Karunaratne, A. and T. Wheeler, 2015: Observed relationships between maize yield and climate in sri lanka. *Agronomy Journal*, **107**(1), 395–405.

Kassie, B., M. Van Ittersum, H. Hengsdijk, S. Asseng, J. Wolf, and R. P. Rötter, 2014: Climate-induced yield variability and yield gaps of maize (zea mays l.) in the central rift valley of ethiopia. *Field Crops Research*, **160**, 41–53.

Kassie, B. T., 2014: *Climate variability and change in Ethiopia: exploring impacts and adaptation options for cereal production.* PhD thesis, Wageningen University,NL.

Kassie, B. T., S. Asseng, R. P. Rotter, H. Hengsdijk, A. C. Ruane, and M. K. Van Ittersum, 2015: Exploring climate change impacts and adaptation options for maize production in the central rift valley of ethiopia using different climate change scenarios and crop models. *Climatic Change*, **129**(1-2), 145–158.

Keita, N. and E. Carfagna, 2010: Overview of methodological issues for research to improve agricultural statistics in developing countries. In *Proceeding of ICAS-V, Fifth International Conference on Agricultural Statistics, Integrating Agriculture into National Statistical Systems Kampala, Uganda*, volume 1, 26.

Keulen, H. v. and J. Wolf, 1986: *Modelling of agricultural production: weather, soils and crops.* Pudoc.

Kharin, V. V. and F. W. Zwiers, 2003: On the roc score of probability forecasts. *Journal of Climate*, **16**(24), 4145–4150.

Kim, H.-M., P. J. Webster, and J. A. Curry, 2012: Seasonal prediction skill of ecmwf system 4 and ncep cfsv2 retrospective forecast for the northern hemisphere winter. *Climate Dynamics*, **39**(12), 2957–2973.

Kim, H.-M., P. J. Webster, J. A. Curry, and V. E. Toma, 2012: Asian summer monsoon prediction in ecmwf system 4 and ncep cfsv2 retrospective seasonal forecasts. *Climate Dynamics*, **39**(12), 2975–2991.

Komarek, A. M., J. Thurlow, J. Koo, and A. De Pinto, 2019: Economywide effects of climate-smart agriculture in ethiopia. *Agricultural Economics*, **50**(6), 765–778.

Korecha, D. and A. G. Barnston, 2007: Predictability of june–september rainfall in ethiopia. *Monthly Weather Review*, **135**(2), 628–650.

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.

Kuhnert, M., J. Yeluripati, P. Smith, H. Hoffmann, M. Van Oijen, J. Constantin, E. Coucheney, R. Dechow, H. Eckersten, T. Gaiser, et al., 2017: Impact analysis of climate data aggregation at different spatial scales on simulated net primary productivity for croplands. *European Journal of Agronomy*, **88**, 41–52.

Kumar, A., A. G. Barnston, and M. P. Hoerling, 2001: Seasonal predictions, probabilistic verifications, and ensemble size. *Journal of Climate*, **14**(7), 1671–1676.

Kumar, K. K., B. Rajagopalan, and M. A. Cane, 1999: On the weakening relationship between the indian monsoon and enso. *Science*, **284**(5423), 2156–2159.

Lafon, T., S. Dadson, G. Buys, and C. Prudhomme, 2013: Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *International Journal of Climatology*, **33**(6), 1367–1381.

Latif, M., D. Dommenget, M. Dima, and A. Grötzner, 1999: The role of indian ocean sea surface temperature in forcing east african rainfall anomalies during december–january 1997/98. *Journal of Climate*, **12**(12), 3497–3504.

Lawless, C. and M. A. Semenov, 2005: Assessing lead-time for predicting wheat growth using a crop simulation model. *Agricultural and Forest Meteorology*, **135**(1-4), 302–313.

Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams, D. B. Wuertz, R. S. Vose, and J. Rennie, 2011: An overview of the global historical climatology network monthly mean temperature data set, version 3. *Journal of Geophysical Research: Atmospheres*, **116**(D19).

Legates, D. R. and G. J. McCabe Jr, 1999: Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, **35**(1), 233–241.

Liebmann, B., M. P. Hoerling, C. Funk, I. Bladé, R. M. Dole, D. Allured, X. Quan, P. Pegion, and J. K. Eischeid, 2014: Understanding recent eastern horn of africa rainfall variability and change. *Journal of Climate*, **27**(23), 8630–8645.

Liu, M., K. Rajagopalan, S. Chung, X. Jiang, J. Harrison, T. Nergui, A. Guenther, C. Miller, J. Reyes, C. Tague, et al., 2014: What is the importance of climate model bias when projecting the impacts of climate change on land surface processes? *Biogeosciences*, **11**(10), 2601–2622.

Liu, W., B. Huang, P. W. Thorne, V. F. Banzon, H.-M. Zhang, E. Freeman, J. Lawrimore, T. C. Peterson, T. M. Smith, and S. D. Woodruff, 2015: Extended reconstructed sea surface temperature version 4 (ersst. v4): Part ii. parametric and structural uncertainty estimations. *Journal of Climate*, **28**(3), 931–951.

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

Lyon, B. and S. J. Mason, 2009: The 1997/98 summer rainfall season in southern africa. part ii: Model simulations and coupled model forecasts. *Journal of Climate*, **22**(13), 3802–3818.

Ma, J. and J.-F. Maystadt, 2017: The impact of weather variations on maize yields and household income: income diversification as adaptation in rural china. *Global Environmental Change*, **42**, 93–106.

Macadam, I., D. Argüeso, J. P. Evans, D. L. Liu, and A. J. Pitman, 2016: The effect of bias correction and climate model resolution on wheat simulations forced with a regional climate model ensemble. *International Journal of Climatology*, **36**(14), 4577–4591.

Makio, J., 2020: Description of cropping systems, climate and soils in tanzania. Avaialable at `http://www.yieldgap.org/kenya`(2020/05/06).

Manzanas, R., M. Frías, A. Cofiño, and J. M. Gutiérrez, 2014: Validation of 40 year multimodel seasonal precipitation forecasts: The role of enso on the global skill. *Journal of Geophysical Research: Atmospheres*, **119**(4), 1708–1719.

Maraun, D., F. Wetterhall, A. Ireson, R. Chandler, E. Kendon, M. Widmann, S. Brienen, H. Rust, T. Sauter, M. Themeßl, et al., 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, **48**(3).

Marletto, V., F. Zinoni, L. Criscuolo, G. Fontana, S. Marchesi, A. Morgillo, M. Van Soetendael, E. Ceotto, and U. Andersen, 2005: Evaluation of downscaled demeter multi-model ensemble seasonal hindcasts in a northern italy location by means of a model of wheat growth and soil water balance. *Tellus A: Dynamic Meteorology and Oceanography*, **57**(3), 488–497.

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

Martinez, R., B. Garanganga, A. Kamga, Y. Luo, S. Mason, J. Pahalad, and M. Rummukainen, 2010: Regional climate information for risk management: capabilities. *Procedia Environmental Sciences*, **1**, 354–368.

Martre, P., D. Wallach, S. Asseng, F. Ewert, J. W. Jones, R. P. Rotter, K. J. Boote, A. C. Ruane, P. J. Thorburn, D. Cammarano, et al., 2015: Multimodel ensembles of wheat growth: More models are better than one.

Mason, S. J., 2004: On using "climatology" as a reference strategy in the brier and ranked probability skill scores. *Monthly Weather Review*, **132**(7), 1891–1895.

Mason, S. J. and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. *Weather and Forecasting*, **14**(5), 713–725.

Mason, S. J. and D. B. Stephenson, 2008: How do we know whether seasonal climate forecasts are any good? In *Seasonal Climate: Forecasting and Managing Risk*. Springer, 259–289.

Mason, S. J. and A. P. Weigel, 2009: A generic forecast verification framework for administrative purposes. *Monthly Weather Review*, **137**(1), 331–349.

Matthew, O. J., B. J. Abiodun, and A. T. Salami, 2015: Modelling the impacts of climate variability on crop yields in nigeria: performance evaluation of regcm3-glam system. *Meteorological Applications*, **22**(2), 198–212.

McIntosh, P. C., A. J. Ash, and M. S. Smith, 2005: From oceans to farms: the value of a novel statistical climate forecast for agricultural management. *Journal of Climate*, **18**(20), 4287–4302.

McIntosh, P. C., M. J. Pook, J. S. Risbey, S. N. Lisson, and M. Rebbeck, 2007: Seasonal climate forecasts for agriculture: Towards better understanding and value. *Field Crops Research*, **104**(1-3), 130–138.

Meehl, G. A., C. A. Senior, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, **9**(5), 1937.

Meyers, G., P. McIntosh, L. Pigot, and M. Pook, 2007: The years of el niño, la niña, and interactions with the tropical indian ocean. *Journal of Climate*, **20**(13), 2872–2880.

Mjelde, J. W. and K. Keplinger, 1998: Using the southern oscillation to forecast texas winter wheat and sorghum crop yields. *Journal of Climate*, **11**(1), 54–60.

Mlingano, 2006: Soils of tanzania and their potential for agriculture development.

Mlynczak, P. E., G. Smith, and P. Stackhouse Jr, 2010: Interannual variations of surface radiation budget. In *22nd Conference on Climate Variability and Change*.

Molteni, F., T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen, T. Palmer, and F. Vitart, 2011: *The new ECMWF seasonal forecast system (System 4)*, volume 49. European Centre for Medium-Range Weather Forecasts Reading.

Monti, A. and G. Venturi, 2007: A simple method to improve the estimation of the relationship between rainfall and crop yield. *Agronomy for Sustainable Development*, **27**(3), 255–260.

Morcrette, J., H. W. Barker, J. Cole, M. J. Iacono, and R. Pincus, 2008: Impact of a new radiation package, mcrad, in the ecmwf integrated forecasting system. *Monthly Weather Review*, **136**(12), 4773–4798.

Morcrette, J.-J., G. Mozdzynski, and M. Leutbecher, 2008: A reduced radiation grid for the ecmwf integrated forecasting system. *Monthly Weather Review*, **136**(12), 4760–4772.

Morse, A. P., F. J. Doblas-Reyes, M. B. Hoshen, R. Hagedorn, and T. N. Palmer, 2005: A forecast quality assessment of an end-to-end probabilistic multi-model seasonal forecast system using a malaria model. *Tellus A: Dynamic Meteorology and Oceanography*, **57**(3), 464–475.

Mtongori, H. I., F. Stordal, R. Benestad, S. Mourice, M. Pereira-Flores, and F. Justino, 2015: Impacts of climate and farming management on maize yield in southerntanzania. *African Crop Science Journal*, **23**(4), 399–417.

Müller, W., C. Appenzeller, F. Doblas-Reyes, and M. Liniger, 2005: A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *Journal of Climate*, **18**(10), 1513–1523.

Murphy, A. H. and D. S. Wilks, 1998: A case study of the use of statistical models in forecast verification: Precipitation probability forecasts. *Weather and Forecasting*, **13**(3), 795–810.

Mutai, C. C. and M. N. Ward, 2000: East african rainfall and the tropical circulation/convection on intraseasonal to interannual timescales. *Journal of Climate*, **13**(22), 3915–3939.

Mwangi, E., F. Wetterhall, E. Dutra, F. Di Giuseppe, and F. Pappenberger, 2014: Forecasting droughts in east africa. *Hydrology and Earth System Sciences*, **18**(2), 611.

Nadler, A. and P. R. Bullock, 2011: Long-term changes in heat and moisture related to corn production on the canadian prairies. *Climatic Change*, **104**(2), 339–352.

Neumann, K., P. H. Verburg, E. Stehfest, and C. Müller, 2010: The yield gap of global grain production: A spatial analysis. *Agricultural Systems*, **103**(5), 316–326.

Nicholson, S. E., 2014: The predictability of rainfall over the greater horn of africa. part i: Prediction of seasonal rainfall. *Journal of Hydrometeorology*, **15**(3), 1011–1027.

Novella, N. S. and W. M. Thiaw, 2013: African rainfall climatology version 2 for famine early warning systems. *Journal of Applied Meteorology and Climatology*, **52**(3), 588–606.

Ogallo, L., P. Bessemoulin, J.-P. Ceron, S. Mason, and S. J. Connor, 2008: Adapting to climate variability and change: the climate outlook forum process. *Bulletin of the World Meteorological Organization*, **57**(2), 93–102.

Ogutu, G. E., W. H. Franssen, I. Supit, P. Omondi, and R. W. Hutjes, 2017: Skill of ecmwf system-4 ensemble seasonal climate forecasts for east africa. *International Journal of Climatology*, **37**(5), 2734–2756.

Ogutu, G. E., W. H. Franssen, I. Supit, P. Omondi, and R. W. Hutjes, 2018: Probabilistic maize yield prediction over east africa using dynamic ensemble seasonal climate forecasts. *Agricultural and Forest Meteorology*, **250**, 243–261.

Ogwang, B. A., H. Chen, X. Li, and C. Gao, 2014: The influence of topography on east african october to december climate: sensitivity experiments with regcm4. *Advances in Meteorology*, **2014**.

Ojeda, J. J., E. E. Rezaei, T. A. Remenyi, M. A. Webb, H. A. Webber, B. Kamali, R. M. Harris, J. N. Brown, D. B. Kidd, C. L. Mohammed, et al., 2020: Effects of soil-and climate data aggregation on simulated potato yield and irrigation water requirement. *Science of The Total Environment*, **710**, 135589.

Olesen, J. E., P. K. Bøcher, and T. Jensen, 2000: Comparison of scales of climate and soil data for aggregating simulated yields of winter wheat in denmark. *Agriculture, Ecosystems & Environment*, **82**(1-3), 213–228.

Omondi, P., L. A. Ogallo, R. Anyah, J. Muthama, and J. Ininda, 2013: Linkages between global sea surface temperatures and decadal rainfall variability over eastern africa region. *International Journal of Climatology*, **33**(8), 2082–2104.

Omoyo, N. N., J. Wakhungu, and S. Oteng'i, 2015: Effects of climate variability on maize yield in the arid and semi arid lands of lower eastern kenya. *Agriculture & Food Security*, **4**(1), 8.

Omuto, C. T., 2013: Major soil and data types in kenya. In *Developments in Earth Surface Processes*, volume 16. Elsevier, 123–132.

Owiti, Z. and L. Ogallo, 2007: Dynamical aspects of indian ocean dipole and east african rainfall anomaly linkages. *J. Meteorol. Relat. Sci*, **1**(1), 14–19.

Owiti, Z., L. A. Ogallo, and J. Mutemi, 2008: Linkages between the indian ocean dipole and east african seasonal rainfall anomalies. *Journal of Kenya Meteorological Society Volume*, **2**(1).

Owiti, Z. and W. Zhu, 2012: Spatial distribution of rainfall seasonality over east africa. *Journal of Geography and Regional Planning*, **5**(15), 409.

Palmer, T. N., 2006: *Predictability of weather and climate: from theory to practice.* Cambridge University Press, 1–29.

Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Delécluse, M. Déqué, E. Diez, F. J. Doblas-Reyes, H. Feddersen, et al., 2004: Development of a european multimodel ensemble system for seasonal-to-interannual prediction (demeter). *Bulletin of the American Meteorological Society*, **85**(6), 853–872.

Parton, K. A., J. Crean, and P. Hayman, 2019: The value of seasonal climate forecasts for australian agriculture. *Agricultural Systems*, **174**, 1–10.

Philippon, N., P. Camberlin, and N. Fauchereau, 2002: Empirical predictability study of october–december east african rainfall. *Quarterly Journal of the Royal Meteorological Society*, **128**(585), 2239–2256.

Phillips, J., M. Cane, and C. Rosenzweig, 1998: Enso, seasonal rainfall patterns and simulated maize yield variability in zimbabwe. *Agricultural and Forest Meteorology*, **90**(1-2), 39–50.

Porter, J. R. and M. A. Semenov, 2005: Crop responses to climatic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**(1463), 2021–2035.

Porwollik, V., C. Müller, J. Elliott, J. Chryssanthacopoulos, T. Iizumi, D. K. Ray, A. C. Ruane, A. Arneth, J. Balkovič, P. Ciais, et al., 2017: Spatial and temporal uncertainty of crop yield aggregations. *European Journal of Agronomy*, **88**, 10–21.

Raschke, E., S. Bakan, and S. Kinne, 2006: An assessment of radiation budget data provided by the isccp and gewex-srb. *Geophysical Research Letters*, **33**(7).

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

Resop, J. P., D. H. Fleisher, Q. Wang, D. J. Timlin, and V. R. Reddy, 2012: Combining explanatory crop models with geospatial data for regional analyses of crop yield using field-scale modeling units. *Computers and Electronics in Agriculture*, **89**, 51–61.

Rodriguez, D., P. De Voil, D. Hudson, J. Brown, P. Hayman, H. Marrou, and H. Meinke, 2018: Predicting optimum crop designs using crop models and seasonal climate forecasts. *Scientific Reports*, **8**(1), 1–13.

Rosenzweig, C., J. W. Jones, J. L. Hatfield, A. C. Ruane, K. J. Boote, P. Thorburn, J. M. Antle, G. C. Nelson, C. Porter, S. Janssen, et al., 2013: The agricultural model intercomparison and improvement project (agmip): protocols and pilot studies. *Agricultural and Forest Meteorology*, **170**, 166–182.

Rotach, M. W. and D. Zardi, 2007: On the boundary-layer structure over highly complex terrain: Key findings from map. *Quarterly Journal of the Royal Meteorological Society*, **133**(625), 937–948.

Rotter, R., 2014: *Simulation of the biophysical limitations to maize production under rainfed conditions in Kenya. Evaluation and application of the model WOFOST.* PhD thesis, Wageningen University and Research Centre, Wageningen University, NL.

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

Saji, N. and T. Yamagata, 2003: Possible impacts of indian ocean dipole mode events on global climate. *Climate Research*, **25**(2), 151–169.

Saxton, K., W. Rawls, J. Romberger, and R. Papendick, 1986: Estimating generalized soil-water characteristics from texture. *Soil Science Society of America journal*, **50**(4), 1031–1036.

Semazzi, F., V. Mehta, and Y. Sud, 1988: An investigation of the relationship between sub-saharan rainfall and global sea surface temperatures. *Atmosphere-Ocean*, **26**(1), 118–138.

Semenov, M. A. and F. J. Doblas-Reyes, 2007: Utility of dynamical seasonal forecasts in predicting crop yield. *Climate Research*, **34**(1), 71–81.

Shibu, M., P. Leffelaar, H. Van Keulen, and P. Aggarwal, 2010: Lintul3, a simulation model for nitrogen-limited situations: Application to rice. *European Journal of Agronomy*, **32**(4), 255–271.

Shin, D., G. Baigorria, Y. Lim, S. Cocke, T. LaRow, J. J. O'brien, and J. W. Jones, 2009: Assessing crop yield simulations with various seasonal climate data. In *Proceedings of the 7th NOAA Annual Climate Prediction Application Science Workshop, Norman, OK.*

Shin, D., G. Baigorria, Y. Lim, S. Cocke, T. LaRow, J. J. O'Brien, and J. W. Jones, 2010: Assessing maize and peanut yield simulations with various seasonal climate data in the southeastern united states. *Journal of Applied Meteorology and Climatology*, **49**(4), 592–603.

Shutts, G., M. Leutbecher, A. Weisheimer, T. Stockdale, L. Isaksen, and M. Bonavita, 2011: Representing model uncertainty: Stochastic parametrizations at ecmwf. *ECMWF Newsletter*, **129**, 19–24.

Siegert, S., 2015: Specsverification: forecast verification routines for the specs fp7 project. *URL http://CRAN. R-project. org/package= SpecsVerification, r package version 0.4-1.*

Skees, J., P. Hazell, and M. Miranda, 1999: New approaches to public/private crop yield insurance. *Unpublished Working Paper, The World Bank, Washington, DC.*

Slingo, J. M., A. J. Challinor, B. J. Hoskins, and T. R. Wheeler, 2005: Introduction: food crops in a changing climate. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**(1463), 1983–1989.

Smith, K. A. and F. H. Semazzi, 2014: The role of the dominant modes of precipitation variability over eastern africa in modulating the hydrology of lake victoria. *Advances in Meteorology*, **2014**.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to noaa's historical merged land–ocean surface temperature analysis (1880–2006). *Journal of Climate*, **21**(10), 2283–2296.

Song, Y., F. H. Semazzi, L. Xie, and L. J. Ogallo, 2004: A coupled regional climate model for the lake victoria basin of east africa. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **24**(1), 57–75.

Srivastava, A. K., C. M. Mboh, T. Gaiser, H. Webber, and F. Ewert, 2016: Effect of sowing date distributions on simulation of maize yields at regional scale–a case study in central ghana, west africa. *Agricultural Systems*, **147**, 10–23.

Stehfest, E., M. Heistermann, J. A. Priess, D. S. Ojima, and J. Alcamo, 2007: Simulation of global crop production with the ecosystem model daycent. *Ecological Modelling*, **209**(2-4), 203–219.

Stockdale, T. N., O. Alves, G. Boer, M. Deque, Y. Ding, A. Kumar, K. Kumar, W. Landman, S. Mason, P. Nobre, et al., 2010: Understanding and predicting seasonal-to-interannual climate variability-the producer perspective. *Procedia Environmental Sciences*, **1**, 55–80.

Stockdale, T. N., D. L. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature*, **392**(6674), 370–373.

Supit, I. and E. Van der Goot, 1999: National wheat yield prediction of france as affected by the prediction level. *Ecological Modelling*, **116**(2-3), 203–223.

Supit, I., C. Van Diepen, A. De Wit, P. Kabat, B. Baruth, and F. Ludwig, 2010: Recent changes in the climatic yield potential of various crops in europe. *Agricultural Systems*, **103**(9), 683–694.

Supit, I., C. Van Diepen, A. De Wit, J. Wolf, P. Kabat, B. Baruth, and F. Ludwig, 2012: Assessing climate change effects on european crop yields using the crop growth monitoring system and a weather generator. *Agricultural and Forest Meteorology*, **164**, 96–111.

Taffesse, A. S., P. Dorosh, and S. A. Gemessa, 2012: Crop production in ethiopia: Regional patterns and trends. *Food and agriculture in Ethiopia: Progress and policy challenges* 53–83.

Tanarhte, M., P. Hadjinicolaou, and J. Lelieveld, 2012: Intercomparison of temperature and precipitation data sets based on observations in the mediterranean and the middle east. *Journal of Geophysical Research: Atmospheres*, **117**(D12).

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

Tebaldi, C., R. L. Smith, D. Nychka, and L. O. Mearns, 2005: Quantifying uncertainty in projections of regional climate change: A bayesian approach to the analysis of multimodel ensembles. *Journal of Climate*, **18**(10), 1524–1540.

Tesfaye, K., M. Kassie, J. E. Cairns, M. Michael, C. Stirling, T. Abate, B. Prasanna, M. Mekuria, H. Hailu, O. Erenstein, et al., 2017: Potential for scaling up climate smart agricultural practices: examples from sub-saharan africa. In *Climate Change Adaptation in Africa*. Springer, 185–203.

Thiery, W., E. L. Davin, H.-J. Panitz, M. Demuzere, S. Lhermitte, and N. Van Lipzig, 2015: The impact of the african great lakes on the regional climate. *Journal of Climate*, **28**(10), 4061–4085.

Thomson, M. C., F. Doblas-Reyes, S. J. Mason, R. Hagedorn, S. J. Connor, T. Phindela, A. Morse, and T. Palmer, 2006: Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, **439**(7076), 576–579.

Thornton, P. K., P. G. Jones, G. Alagarswamy, and J. Andresen, 2009: Spatial variation of crop yield response to climate change in east africa. *Global Environmental Change*, **19**(1), 54–65.

Tippett, M. K. and A. G. Barnston, 2008: Skill of multimodel enso probability forecasts. *Monthly Weather Review*, **136**(10), 3933–3946.

Troccoli, A., M. Harrison, M. Coughlan, and J. B. Williams, 2008: Seasonal forecasts in decision making. In *Seasonal climate: forecasting and managing risk*. Springer, 13–41.

Ubilava, D. and M. Abdolrahimi, 2019: The el niño impact on maize yields is amplified in lower income teleconnected countries. *Environmental Research Letters*, **14**(5), 054008.

Uppala, S. M., P. Kållberg, A. Simmons, U. Andrae, V. D. C. Bechtold, M. Fiorino, J. Gibson, J. Haseler, A. Hernandez, G. Kelly, et al., 2005: The era-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131**(612), 2961–3012.

Van Bussel, L., F. Ewert, and P. Leffelaar, 2011: Effects of data aggregation on simulations of crop phenology. *Agriculture, Ecosystems & Environment*, **142**(1-2), 75–84.

Van Den Dool, H. M. and Z. Toth, 1991: Why do forecasts for "near normal" often fail? *Weather and Forecasting*, **6**(1), 76–85.

Van Diepen, C., C. Rappoldt, J. Wolf, and H. Van Keulen, 1988: Crop growth simulation model wofost version 4.1, documentation. In *SOW-88-01*. Centre for World Food Studies Wageningen, 299.

Van Diepen, C. v., J. Wolf, H. Van Keulen, and C. Rappoldt, 1989: Wofost: a simulation model of crop production. *Soil Use and Management*, **5**(1), 16–24.

Van Heemst, H., 1988: *Plant data values required for simple crop growth simulation models: review and bibliography*. Number 17. CABO [etc.].

van Loon, M. P., N. Deng, P. Grassini, J. I. R. Edreira, E. Wolde-Meskel, F. Baijukya, H. Marrou, and M. K. van Ittersum, 2018: Prospect for increasing grain legume crop production in east africa. *European Journal of Agronomy*, **101**, 140–148.

Ververs, M.-T., 2012: The east african food crisis: did regional early warning systems function? *The Journal of nutrition*, **142**(1), 131–133.

Vision, K., 2007: 2030–the popular version. *Government of Kenya, Nairobi*.

Walker, D. P., C. E. Birch, J. H. Marsham, A. A. Scaife, R. J. Graham, and Z. T. Segele, 2019: Skill of dynamical and ghacof consensus seasonal forecasts of east african rainfall. *Climate Dynamics*, **53**(7-8), 4911–4935.

Weedon, G., P. s. J. A. N. B. O. B. Gomes, S Viterbo, and M. Best, 2010: The watch forcing data 1958–2001: A meteorological forcing dataset for land surface- and hydrological models. Technical report, WATCH Tech. Rep.

Weedon, G., S. Gomes, P. Viterbo, W. J. Shuttleworth, E. Blyth, H. Österle, J. Adam, N. Bellouin, O. Boucher, and M. Best, 2011: Creation of the watch forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *Journal of Hydrometeorology*, **12**(5), 823–848.

Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo, 2014: The wfdei meteorological forcing data set: Watch forcing data methodology applied to era-interim reanalysis data. *Water Resources Research*, **50**(9), 7505–7514.

Weigel, A. P., M. Liniger, and C. Appenzeller, 2008: Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society*, **134**(630), 241–260.

Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007a: The discrete brier and ranked probability skill scores. *Monthly Weather Review*, **135**(1), 118–124.

Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007b: Generalization of the discrete brier and ranked probability skill scores for weighted multimodel ensemble forecasts. *Monthly Weather Review*, **135**(7), 2778–2785.

Weigel, A. P. and S. J. Mason, 2011: The generalized discrimination score for ensemble forecasts. *Monthly Weather Review*, **139**(9), 3069–3074.

Weisheimer, A. and T. Palmer, 2014: On the reliability of seasonal climate forecasts. *Journal of the Royal Society Interface*, **11**(96), 20131162.

Weisheimer, A., T. Palmer, and F. Doblas-Reyes, 2011: Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophysical Research Letters*, **38**(16).

Willmott, C. J., 1981: On the validation of models. *Physical Geography*, **2**(2), 184–194.

Willmott, C. J., 1982: Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, **63**(11), 1309–1313.

Willmott, C. J. and K. Matsuura, 2001: Terrestrial air temperature and precipitation: Monthly and annual time series (1950–1999) version 1.02. *Center for Climatic Research, University of Delaware, Newark*.

Willmott, C. J. and K. Matsuura, 2005: Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate Research*, **30**(1), 79–82.

Willmott, C. J., S. M. Robeson, and K. Matsuura, 2012: A refined index of model performance. *International Journal of Climatology*, **32**(13), 2088–2094.

Wolf, J., K. Ouattara, and I. Supit, 2015: Sowing rules for estimating rainfed yield potential of sorghum and maize in burkina faso. *Agricultural and Forest Meteorology*, **214**, 208–218.

World-Bank Tanzania economic update. Technical report.

Zhang, T., P. W. Stackhouse Jr, S. K. Gupta, S. J. Cox, J. C. Mikovitz, and L. M. Hinkelman, 2013: The validation of the gewex srb surface shortwave flux data products using bsrn measurements: A systematic quality control, production and application approach. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **122**, 127–140.

Zhang, Y., X. Li, and Y. Bai, 2015: An integrated approach to estimate shortwave solar radiation on clear-sky days in rugged terrain using modis atmospheric products. *Solar Energy*, **113**, 347–357.

Zhao, J., J. Guo, and J. Mu, 2015: Exploring the relationships between climatic variables and climate-induced yield of spring maize in northeast china. *Agriculture, Ecosystems & Environment*, **207**, 79–90.

Summary

Climate variability is an important driver for regional anomalous production levels of especially rainfed crops, with implication for food security of subsistence farmers and economic performance for market oriented agriculturalists. In large parts of the tropics modern seasonal ensemble forecast systems have useful levels of skill, that open up the possibility to develop climate services that assist agriculturalist and others in the food chain (farm suppliers, commodity traders, and aid organizations) to anticipate on expected anomalous conditions of an approaching season. In this thesis we explore the forecast skill at various steps in the climate-crop modelling chain for seasonal maize yield anomalies in East Africa. The first part of the thesis was to set up a crop model meant for use in farm scales for regional yield simulation (chapter 2). The main aim was establishing skill of climate forecasts (chapter 3), these provided input to the rest of thesis chapters (i.e. chapter 4, 5 and 6). Besides climate data, other data sets that include land use maps, soil data, crop calendars, and crop parameters are required for the crop model. Some of the data sets are of coarse resolution yet the crop model (WOFOST) model requires homogeneous inputs for simulation. These were therefore overlayed with high resolution land use data to derive simulation units at 0.1° grids that met assumptions of homogeneity in each grid cell. Similarly attributed to the grids are observed and forecast climate data. Planting dates for each grid cell was fixed to dates that resulted in highest average water-limited yields. Harvest dates were not fixed, neither in the baseline nor reforecast yields. The baseline and reforecast harvest dates had a maximum difference of 10-days.

Next, we assessed the skill of ECMWF System-4 seasonal climate forecasts for primary meteorological variables needed for crop modelling against a number of gridded observations. A number of observation datasets that had no dependency with each other were used to ascertain the robustness of the results. We found both potential and real skill for rainfall and temperature in typical cropping seasons in eastern Africa. The skill of the three variables (*tp*, *tas* and *rsds*) depends on season, forecast lead-time and location. Climate forecasts realistically reproduces the spatial patterns of observed seasonal climatology irrespective of the observation data sets considered but with biases. Precipitation forecast biases are season dependent and magnitude depend on the validating observation data for example, precipitation forecasts show dry bias in March-May season, wet biases on October to December season, but both dry and wet biases in certain regions in June-August season. Cold biases dominate

tas forecasts in all seasons, the spatial patterns have no strong seasonal variation. Patterns in downward surface shortwave radiation depend on prevailing rain seasons and the surface topography. The ROCSS reveals that *S4* skillfully discriminates between upper-tercile, middle-tercile and lower-tercile forecasts. Aggregated over homogeneous rainfall regions of East Africa, upper-tercile and lower-tercile precipitation forecasts initialized up to at least three lead-months show skill. Temperature forecasts initialized up to 4 months before start of season are skillful and useful while *rsds* show less skill though still usable. Middle-tercile predictions remain poor for all variables at all seasons and lead-times. But forecast skill and lead-times of useful forecasts are season and region specific. We investigate the influence of bias correction in forecast skill but found no improvement in probabilistic forecast skill. Bias correction is however still important for climate impact applications. In general, the forecast fairly captures the inter-annual climate variability of the three variables and may potentially be used to predict impacts.

Next, we analyze correlations between reported production and anomalous weather conditions, using a range of climate indicators relevant for arable farming such as growing and killing degree days, rainfall amount, evenness, random independent events (unevenness), and timing during consequent maize growth phases in two case study regions. This was done for two case study regions i.e. northern Ethiopia and an equatorial region in Kenya. But sensitivities to the weather is not uniform as they affect yields differently. Regionally sensitivity to rainfall is found for northern Ethiopia while a sensitivity to temperature is found in equatorial west Kenya. Some of the variables are predictable while others are not. But in general, the predictability depends on region, maize phenological stage, and climate variable. In this part of the study, significant levels of correlation and skill are revealed that open up the potential for statistical forecasting. Sensitivities to climate indicators at different stages of crop growth provide the possibility of developing mitigation measures that are short term and targeted to particular crop growth phases.

At the next level of complexity we explore the use of full process based crop models forced by seasonal climate forecasts to forecast anomalous water-limited maize yield in the region. We generated a 15-member ensemble of yield predictions using the World Food Studies (WOFOST) crop model implemented for water-limited maize production and single season simulation. Maize yield predictions are validated against baseline yield, also simulated from driving the crop model with historical observations. Focusing on the dominant sowing dates in the northern region (July), equatorial region (March-April) and in the southern region (December), we find again potentially useful levels of skill with at least two months lead before planting, in most agricultural regions. The baseline yield anomalies also showed good anomaly correlations compared to the official FAO and national reported statistics, but the average reference yield values are lower than those reported in Kenya and Ethiopia, but slightly higher in Tanzania. Notable differences in magnitude of yields may be attributed to number of crop seasons i.e. some regions have double crop seasons yet the crop model was set up to simulate only a single season. Difference in interannual variability between the baseline and predicted yields range from from ±40%, but higher interannual variability in predicted yield dominates. Anomaly correlations between the reference and predicted yields are largely positive and range from +0.3 to +0.6. Above-normal and below-normal yield anomalies are predictable with at least 2-months lead-time in many agricultural regions. In areas where yield forecast skill exist, it is possible to provide quantitative crop yield outlooks by combining dynamic crop yield forecasts and crop models.

Finally, we try to attribute skill levels to physiographic characteristics (soils, maize cultivars, geographical region etc.) and address some issues of scale of aggregation for two case study regions in Kenya and Ethiopia. Skill in aggregated yields are driven by physiography, and size of spatial aggregation unit, i.e. better skill when yields are aggregated over a larger area. This could be related to the scale at which physiographic features of an area are captured in the driving climate data. Comparable level of skill is observed when yields are aggregated by soil types. Good, significant prediction skill in individual high resolution grid cells are degraded little when averaged over sub-national boundaries, but then, skill in individual cells are important only if neighbouring cells also have skill. Observed yield statistics are available at national or sub-national regions, skill in aggregated data to these boundaries mean that they could still be useful for regional assessments. But good predictability in individual high resolution cells could still be useful to farmers. Also, high resolution climate data may improve simulation. This may be possible in future as resolutions of climate models increasing get higher.

Samenvatting

Naar seizoensvoorspellingen van maisopbrengsten in Oost Afrika: voorspelbaarheid in de modelketen als basis voor klimaatdiensten voor de landbouw.

Klimaatvariabiliteit speelt een belangrijke rol bij regionale gewas productie anomalieën van de regenafhankelijke landbouw. Voor de voedselzekerheid van zelfvoorzienende boeren evenals voor de economische prestaties voor markt georiënteerde landbouwers kan dit kan gevolgen hebben. In grote delen van de tropen zijn moderne, op ensembles gebaseerde seizoensvoorspellingen van het klimaat beschikbaar. Deze systemen zijn voldoende betrouwbaar om klimaatdiensten te ontwikkelen die landbouwkundigen en anderen in de voedselproductieketen (leveranciers van landbouwproducten, handelaren in grondstoffen, hulporganisaties) kunnen helpen om op verwachte productie afwijkingen van een komend seizoen te anticiperen. In dit proefschrift onderzoeken we de voorspellende kracht in verschillende stappen van de klimaat-gewas modelketen van een dergelijk systeem voor seizoensgebonden maïsopbrengsten in Oost-Afrika. Het eerste deel van dit proefschrift behandelt de opzet van het modelsysteem dat gebruikt is voor opbrengstsimulatie op zowel veldschaal als ook op regionaal niveau (hoofdstuk 2). Het belangrijkste doel van het volgend hoofdstuk is het vaststellen van de nauwkeurigheid en betrouwbaarheid van de klimaatvoorspellingen zelf (hoofdstuk 3). Deze resultaten vormen de basis voor de volgende hoofdstukken van dit proefschrift, waar geanalyseerd wordt in welke mate voorspelbaarheid van het klimaat zich vertaald in voorspelbaarheid van maisopbrengsten, enerzijds op statistische basis (hoofdstuk4), anderzijds met behulp van het in hoofdstuk 2 beschreven gewasgroeimodel (hoofdstuk 5) en wat de rol is van aggregatie in die laatste (hoofdstuk 6). Tot slot volgt een synthese en discussie van de implicaties van de resultaten voor de ontwikkeling van op landbouw gerichte klimaatdiensten in Oost Afrika (hoofdstuk 7).

Naast klimaatgegevens zijn voor het gewasgroeimodel (WOFOST) ook andere datasetsets gebruikt, waaronder landgebruikskaarten, bodemgegevens, gewaskalenders en gewasparameters. De resolutie van deze datasets zijn niet gelijk, sommige hadden een grove resolutie andere een fijnere. Om een homogene invoer voor het model te verkrijgen werd met behulp van een landgebruiksraster simulatie-eenheden met een resolutie van ongeveer 0.1° afgeleid, die voldeden aan de veronderstellingen van homogeniteit in elke rastercel. De waargenomen en voorspelde klimaatgegevens zijn op een dezelfde manier aan de rastercellen toegekend. Gewas-, cq mais variëteiten, gedefinieerd in termen van warmtebehoefte voor bloei en afrijping en plantdatum, werden geoptimaliseerd naar hoogste gemiddelde (waterbeperkte) opbrengsten. De oogstdatums in de basis situatie en de voorspellingen zijn niet opgelegd maar berekend. Deze datums hebben in de basis situatie en in de voorspellingen een verschil van maximaal 10 dagen.

Vervolgens vergelijken we de meteorologische variabelen die nodig zijn voor simulatie van gewasgroei van de ECMWF System-4 historische seizoensvoorspellingen, met een aantal gerasterde waarnemingen en analyseerden we hoe goed deze voorspeld worden. Om de robuustheid van de resultaten te bepalen zijn een aantal onafhankelijke observatiegegevenssets gebruikt. Voor regenval en temperatuur vertoont System-4 in de drie typische teeltseizoenen in Oost-Afrika zowel potentiële als reëel voorspellend vermogen. Het voorspellend vermogen voor de drie variabelen (neerslag *tp*, temperatuur *tas* en zonneschijn *rsds*) hangt af van het seizoen, de voorspeltermijn en de locatie. Klimaatvoorspellingen reproduceren realistische ruimtelijke patronen van de geobserveerde seizoensklimatologie, maar er zijn systematische vertekeningen. Zo vertonen neerslagvoorspellingen een droge bias in de periode maart-mei, een natte bias in de maanden oktober tot december, maar in bepaalde regio's komen in de periode juni-augustus zowel een droge als een natte bias voor. In alle seizoenen overheerst een koude bias de *tas* voorspellingen, de ruimtelijke patronen laten geen sterke seizoensvariatie zien. Patronen in de inkomende kortgolvige straling zijn afhankelijk van de heersende regenseizoenen en de topografie. De ROCSS test laat zien dat *S4* een goed onderscheid maakt tussen voorspellingen van het bovenste terciel, middelste terciel en laagste terciel. Geaggregeerd over homogene regenzones in Oost-Afrika, hebben hoogste terciel en laagste terciel neerslagvoorspellingen die tot en met drie maanden voorafgaand het seizoen gemaakt zijn voorspellende waarde. Temperatuurvoorspellingen die tot 4 maanden voor het begin van het seizoen zijn gemaakt, hebben voorspellende waarde, stralingsvoorspellingen *rsds* hebben minder voorspellende waarde maar zijn nog steeds bruikbaar is. De voorspellingen voor het middelste terciel zijn voor alle variabelen in alle seizoenen en voorspeltermijnen minder bruikbaar. We onderzochten ook de invloed van bias-correctie op de voorspellende waarde, maar vonden geen verbetering in de probabilistische voorspellingen. Bias-correctie is echter nog steeds belangrijk voor toepassingen met betrekking tot klimaat effecten. Over het algemeen geven de voorspellingen een betrouwbare weergave van de klimaatvariabiliteit van de drie variabelen en dit biedt dus mogelijkheden om effecten te voorspellen, op landbouw en/of andere sectoren.

Vervolgens analyseerden we de correlaties tussen de gerapporteerde opbrengsten en afwijkende weersomstandigheden. Bij deze analyse is gebruik gemaakt van een reeks klimaatindicatoren die relevant zijn voor akkerbouw, zoals groei en sterfte graaddagen, eenvoudige temperatuur extremen, de hoeveelheid regen en de temporele spreiding daar van over het groeiseizoen (gelijkmatig of niet, vroeg of laat) tijdens drie opeenvolgende groeifases van maïs (vegetatieve-, bloei- en reproductie fasen). Dit is gedaan voor twee casestudyregio's, namelijk Noord-Ethiopië en een equatoriale regio in Kenia. Weersinvloeden op het gewas zijn niet uniform en hebben in verschillende regio's andere invloeden op de opbrengsten. Regionale gevoeligheid voor regenval wordt gevonden voor Noord-Ethiopië, terwijl temperatuurgevoeligheid wordt gevonden in equatoriaal West-Kenia. Sommige variabelen zijn voorspelbaar, andere niet, maar in het algemeen hangt de voorspelbaarheid af van de regio, groeifase van maïs en klimaatvariabele. In dit deel van de studie werden significante correlaties gevonden die de mogelijkheid voor statistische prognoses openen. Gevoeligheden voor klimaatindicatoren in verschillende stadia van gewasgroei bieden de mogelijkheid om op korte termijn mitigerende maatregelen te ontwikkelen die gericht zijn op bepaalde groeifasen van gewassen.

Op het volgende complexiteit niveau verkenden we het gebruik van een proces georiënteerd gewasgroeimodel met seizoenprognoses als invoer om afwijkingen in de normale waterbeperkte maïsproductie in de regio te voorspellen. We gebruikten een 15-ledig ensemble van waterbeperkte maïsproductie opbrengstvoorspellingen. De opbrengstvoorspellingen zijn gegenereerd met behulp van het World Food Studies (WOFOST) gewasgroeimodel. De voorspelde maïsopbrengst zijn gevalideerd met behulp van de opbrengsten die gesimuleerd zijn met het gewasgroeimodel met historische waarnemingen als input. De afwijkingen in de basis simulatie (historische gewasproductie) vertoonden goede anomaliecorrelaties vergeleken met de officiële FAO en nationaal gerapporteerde statistieken, maar de gemiddelde referentieopbrengsten zijn lager dan die gerapporteerde opbrengsten in Kenia en Ethiopië, maar iets hoger in Tanzania. Opmerkelijke verschillen in de opbrengst omvang kunnen worden toegeschreven aan het aantal oogstseizoenen, dat wil zeggen dat sommige regio's een dubbel groeiseizoen hebben. In de huidige opzet is maar één, het langste seizoen, in beschouwing genomen. Het verschil in variabiliteit in opbrengsten tussen de jaren in de basis simulatie en in de voorspellingen varieert met $\pm 40\%$. Een hogere opbrengst variabiliteit tussen de jaren in de voorspellingen komt vaker voor. Gebruikmakend van de dominante zaaidatums in de noordelijke regio (juli), equatoriale regio (maart-april) en in de zuidelijke regio (december), vinden we opnieuw in de meeste landbouw regio's potentieel bruikbare voorspellingen die ten minste twee maanden voor aanvang van het planten zijn gemaakt. Anomaliecorrelaties tussen de referentie- en voorspelde opbrengsten zijn grotendeels positief en variëren van +0,3 tot +0,6. Afwijkingen boven de normale en onder de normale opbrengst zijn in veel landbouwregio's voorspelbaar met een doorlooptijd van ten minste 2 maanden. In gebieden waar de seizoen voorspellingen voorspellende kracht hebben, is het mogelijk om kwantitatieve gewasopbrengst vooruitzichten te geven door seizoensgebonden klimaatvoorspellingen en een gewasgroeimodel te combineren.

Ten slotte probeerden we de voorspellend vermogen te relateren aan fysiografische kenmerken (bodem, cultivars van maïs, geografische regio, enz.) En behandelden we enkele aggregatie effecten voor twee casestudyregio's in Kenia en Ethiopië. Het voorspellend vermogen van geaggregeerde opbrengst voorspellingen wordt bepaald door fysiografie en de grootte van de ruimtelijke aggregatie-eenheid, d.w.z. betere voorspellingen wanneer opbrengsten over een groter gebied worden geaggregeerd. Dit kan verband houden met de schaal waarop fysiografische kenmerken van een gebied zijn vastgelegd ten opzichte van die van de klimaatgegevens die als invoer gebruikt zijn. Vergelijkbaar voorspellend vermogen wordt waargenomen wanneer opbrengsten naar grondsoorten worden geaggregeerd. Goed, significante voorspellend vermogen van individuele rastercellen met hoge resolutie vermindert weinig wanneer ze gemiddeld worden over sub-nationale grenzen, maar de voorspellende kracht van individuele cellen is dan alleen belangrijk als naburige cellen ook voorspellend vermogen hebben. Opbrengststatistieken zijn beschikbaar op nationaal en regionaal niveau, aggregatie tot dit niveau zou nuttig kunnen zijn wanneer er voorspellend vermogen op dit niveau bestaat. Goede voorspelbaarheid voor individuele hoge resolutie cellen kan nog steeds nuttig zijn voor boeren. Ook kunnen hoge resolutie klimaatgegevens simulaties verbeteren. Naarmate de resoluties van toekomstige klimaatmodellen hoger worden zou dit in de toekomst mogelijk kunnen zijn.

Op basis van voorgaande is aangetoond dat er in de verschillende klimaatzones van Oost Afrika enerzijds voldoende correlaties bestaan tussen inter-jaarlijkse klimaatanomalieën en variaties in maisproductie, en dat er daarnaast voldoende voorspelbaarheid gerealiseerd kan worden in de verschillende stappen van de modelketen, om ruim voor aanvang van het groeiseizoen daarop te kunnen anticiperen met passende adaptatiemaatregelen. We zien hierin een sterk aanmoediging voor verdere ontwikkeling van operationele klimaatdiensten gericht op de landbouwsector in deze regio.

Acknowledgements

First and foremost, I wish to sincerely thank my promoter Dr. Ronald Hutjes for having the confidence to grant this research opportunity. Ronald, thank you for your understanding, support, encouragement throughout the PhD cycle. At times, I felt like giving up but you motivated me to continue the research. Thank you for critical reviews of my work, your dedication to detail and for always being available for meetings and discussions despite your busy schedule. My co-promoter Dr. Iwan Supit, this study would not have been successful without your support especially on agronomy, crop modeling, and SQL database management. Your technical insights, quite critical reviews, and ability to crack jokes even under stressful conditions was motivating. But I cannot also forget how you disapproved poor writing and vocabulary. Thanks to you, it made me strive to always have a "better" first draft of my chapters. My other co-promoter Dr. Philip Omondi, thank you for the encouragement and always thoughtful, critical review of my work. Thank you for being available for consultations during my PhD research.

This study would not have been successful without the support of other people apart from my supervision team. Wietse Franssen, "R" statistical and programming software was alien to me when I started this PhD journey but you helped me through 'baby steps'. I could not have learnt this language without your help. Thank you for always making time to promptly assist when I got stuck even on weekends. Henk van der Ham, thank you for always being available, at times on very short notice to assist with the EUPORIAS Oracle database whenever I got stuck. Prof. Dr. Carolien Kroeze, thank you for always being concerned of Water Systems and Global Change (WSG) Group PhD candidates' welfare, your encouragement, and keeping your office always open for consultations. The friendliness of WSG staff made the research journey more bearable.

The Kenyan community in Wageningen, and by extension the community of East African students made Wageningen stay feel like home away from home. My friend Dr. Robert Ochieng' Mugabe, thanks for your friendship. I finally made it to graduation. I thank the Government of Kenya through Director; Kenya Meteorological Department for granting me the opportunity to pursue this PhD research. My late boss Dr. Samwel Marigi (may his soul R.I.P.), i finally made it, I appreciate that you always showed concern, provided support, and encouragement.

It would have been very difficult to go through the PhD journey without support from my family. My father Mr. Shadrack Ogutu Abang', thank you for your unending support, thank you for being there for my teenage son and taking the role of a father once more. You always made sure my son was comfortable in school while I was away in the Netherlands. My mother and the rest of the family, thank you for your support and for being such a wonderful family. Lastly, I wish to sincerely appreciate my spouse Janet Otieno and my children Edmond and Winfrey. Thank you for bearing long periods without my presence with no complaints, but encouragement. Janet, thank you for taking the role of a father and mother in my absence. Thank you all!

Curriculum Vitae

Geoffrey Ogutu was born in 28th of February 1974 in Siaya, Kenya. After finishing secondary school in 1991, he joined the Institute of Meteorological Training and Research (IMTR), Nairobi in 1992 to undertake a Basic training course in meteorology. Between 1994-2000, he undertook a training course in Electrical/Electronics Engineering at the then Kenya Polytechnic (now Technical University of Kenya) graduating with a Diploma and Higher National Diploma. He worked with the Kenya Meteorological Department (KMD) as a Telecommunication Engineering technician. In the year 2004 he joined the University of Nairobi for a degree course, graduating with Bachelor of Science (BSc) degree in Meteorology in 2008. This marked a career change for Geoffrey who upon completing a post-graduate Meteorologists Operational Training Course (OTC) at IMTR, was deployed as a meteorologist in the Kenya Meteorological Department. In 2010, he got a Netherlands Fellowship Program (NFP-NUFFIC) scholarship to study Master of Science degree in Meteorology and Air Quality (MAQ) in Wageningen University, Netherlands graduating with MSc Meteorology and Air-Quality with a specialization in Climate Change in 2012. He was thereafter deployed to Climate Change division of Kenya Meteorological Department where he was working with mainly a Regional Climate Model. In 2013, he again got a scholarship for Sandwich PhD study in Wageningen University under the project; European Provision of Regional Impacts Assessments on Seasonal and Decadal Timescales (EUPORIAS). After a short visit to Wageningen in 2013 to write his research proposal, he was back to work in KMD as he continued with PhD research, with a back and forth travel to the university in Netherlands. Geoffrey continues to work in the Climate Services Branch of KMD where his role is to analyze past climate and future projections as a basis to provide climate services to Kenyan public.

Publications

Ogutu, G. E., Franssen, W. H., Supit, I., Omondi, P., Hutjes, R. W. (2017). Skill of ECMWF system-4 ensemble seasonal climate forecasts for East Africa. *International Journal of Climatology*, 37(5), 2734-2756.

Ogutu, G. E., Franssen, W. H., Supit, I., Omondi, P., Hutjes, R. W. (2018). Probabilistic maize yield prediction over East Africa using dynamic ensemble seasonal climate forecasts. *Agricultural and Forest Meteorology*, 250, 243-261.

Ogutu, G., Franssen, W., Supit, I., Omondi, P., Hutjes, R., 2016a. Usefulness of ECMWF system-4 ensemble seasonal climate forecasts for East Africa, in *Geophysical Research Abstracts EGU General Assembly*. p. 16370.

Ogutu, G., Supit, I., Hutjes, R., 2016b. Probabilistic maize yield simulation over East Africa using ensemble seasonal climate forecasts, in *Geophysical Research Abstracts EGU General Assembly*. pp. 2016–17111.

Ogutu G.E., Supit I, Omondi P., Hutjes R.W., 2020: Sensitivity and Predictability of Rain-fed Maize Yields to Growing Season Climate Indices: East Africa. *Submitted*

Ogutu G.E.,I. Supit,P. Omondi and R.W. Hutjes, 2020: Spatial Aggregation Units' Influence on Predictability of Seasonal Forecasts of Maize Production: a case study. *submitted*

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

for specialised PhD training

The Netherlands research school for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Geoffrey Evans Owino Ogutu

born on 28 February 1974 in Siaya, Kenya

has successfully fulfilled all requirements of the
educational PhD programme of SENSE.

Wageningen, 14 September 2020

Chair of the SENSE board

Prof. dr. Martin Wassen

The SENSE Director

Prof. Philipp Pattberg

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)

K O N I N K L I J K E N E D E R L A N D S E
A K A D E M I E V A N W E T E N S C H A P P E N

The SENSE Research School declares that **Geoffrey Evans Owino Ogutu** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a work load of 32.7 EC, including the following activities:

SENSE PhD Courses

- Environmental research in context (2015)
- Research in context activity: 'Participating in organizing committee and assembling booklet contributing to Symposium on Water Stressed Futures: Water Systems and Global Change (Wageningen, 15 March 2017)'

Other PhD and Advanced MSc Courses

- The art of modelling, PE&RC and WIMEK (2015)
- Systematic approaches to reviewing literature, Wageningen School of Social Sciences (2015)
- Second EUPORIAS climate services masterclass: water, health and food security in a changing climate, Eurac Research (2016)
- Reviewing a scientific paper, Wageningen Graduate Schools (2015)

External training at a foreign research institute

- R-programming (online), Coursera/John Hopkins University (2016)
- The data scientist's toolbox (online), Coursera/John Hopkins University (2016)
- Introduction to Python (online), Datacamp (2016)

Selection of Poster Presentations

- Probabilistic maize yield simulation over east Africa using ensemble seasonal climate forecasts, EGU conference 19 April 2016, Vienna, Austria
- Usefulness of ECMWF system-4 ensemble seasonal climate forecasts for east Africa, EGU CONFERENCE, 22 April 2016, Vienna, Austria
- Skill assessment of GCM bias corrected seasonal climate forecasts over east Africa. EUPORIAS 2015, 29 September -1 October, 2015, Winterthur, Switzerland

Oral Presentations

- *Pre-season maize yield prediction over eastern Africa*, Greater Horn of Africa Climate Outlook Forum, 29-30 August 2016, Kampala, Uganda
- *Usefulness of ECMWF system-4 ensemble prediction system forecasts for impact assessments*, KMD scientific seminar, 1 July 2016, Nairobi, Kenya

SENSE coordinator PhD education

Dr. ir. Peter Vermeulen

Funding

The research described in this thesis was financially supported by the EUPORIAS project (EUropean Provision of Regional Impact Assessment on Seasonal-to-decadal timescale); grant agreement No. 308291, funded by the European Commission (EU) project in the Seventh Framework Programme through Wageningen University and Research. The Government of Kenya through my employer; The Kenya Meteorological Department supported this research by paying a salary as I continued with research in my home country.
Printing of this thesis report was financed by Wageningen University