

# Relationship of Weather and Maize Yields in Kenya

Monika Novackova, Pedram Rowhani, Martin Todd, Annemie Maertens

*Department of Geography, University of Sussex, Falmer, UK*

October 2018

## Abstract

..the full abstract to be written.. This paper contributes to better understanding of effects of drought on food security in Kenya which should lead to improving of early warning systems and food security. Our dataset consists of yearly panel of 47 counties of Kenya describing the period of 1981-2017.

...Applying the linear mixed effects models, we found that...

**Keywords:** Drought, Early warning systems, Food security, Food systems, Kenya, Mixed effects models

# 1 Introduction and literature review

Consequences of droughts and extreme heat waves on agriculture have been well documented (?????). Agricultural drought and soil moisture are important factors influencing plant health but they are also essential for land-atmospheric feedback and for temperature variability (?).

Frequent occurrence of severe droughts have been a major problem in Kenya and in other Sub-Saharan countries (???). The agricultural sector is the main contributor to the economy of Kenya, hence the effects of drought are especially damaging in this country (?). Besides compromising food security, drought usually leads to severe damage in water supply, electricity and the environment. In 2010-2011, the most devastating drought in decades resulted in food crisis in many East African countries including Kenya (?). Widespread catastrophic droughts occurred also in 1984-1985, 2005 and 2008 (????). As a consequence of climate change, the situation is likely to get worse. Dry areas have strong tendency to get drier while wet areas are getting wetter (???). Strong downwards trend in precipitation has already been observed in the tropics from 10°N to 10°S, especially after 1977 (?). During the period 1900 – 2005, the climate has become wetter in many parts of the world (eastern parts of America, northern Europe, northern and central Asia) but it has become much drier in Mediterranean, Sahel, southern Africa and parts of Southern Asia. Furthermore, increased frequency of heavy rain events has been observed in the areas with decline in total rainfall (?).

Droughts and climate variability do not only affect agriculture but they have widespread consequences on food security in general. ? has found a connection between commodity prices and phases of the ENSO cycle. ? have used a Computable General Equilibrium (CGE) model for Ethiopia and they have shown that production shocks and resulting price

increases have negative effects on farmers' income in drought-prone areas while moderate price increases can be beneficial for farmers in less drought-prone areas. ? have focused on sub-Saharan Africa, central America, North Africa and other low-income countries selected by Oxfam and he has concluded that climate change will result in decline in food productivity while food prices will increase. ? have investigated food prices in 554 local commodity markets in Africa, South Asia and Latin America and they concluded that almost 20% of local market food prices had been affected by weather disturbances.

A large literature has focused on relationship of agricultural production and climate in Sub-Saharan Africa. Increase in temperature has typically been found to affect crop revenue negatively while increase in precipitation has mostly been associated with positive effects in the literature. ? and ? have found these results by the means of Ricardian analysis. ? have focused on Kenya while ? addressed the situation in Niger.

Other studies have investigated the relationship of climate and yields in the context of climate change. ? have developed a simulation model to evaluate impacts of climate change on African agriculture. According to their study, agricultural GDP and climate change exhibit a hill-shaped relationship. Hence, global warming should be beneficial in high latitudes but harmful in low latitudes. The estimated loss of the Africa as a continent ranges from \$25 billion to \$194 billion per year with 2 °C warming (?). ? have shown that in Kenya, higher summer temperatures are harmful while higher winter temperatures are beneficial. Furthermore, fall and summer precipitation is positively correlated with net revenue (?). ? and ? have investigated the distribution of climate change impacts on agriculture in Africa. They argued that mild scenario would result in income gains for African farmers while more severe scenario is likely to be harmful. ? have analysed relationship between climate variability and crop production in Tanzania. The authors have shown that by 2050 the climate change will result in significant decrease in crop

production and big part of this loss can be attributed to change in intra-seasonal variability of weather. ? have estimated effects of climate change on food security in Kenya. They have found that temperature and maize yield exhibit a U-shaped relationship while the relationship of rainfall and maize is hill-shaped. Thus, sufficient precipitation is crucial, but excessive rainfall is harmful for maize yields. ? has reviewed several studies on effects of climate change on agriculture in developing countries. According to the literature review, the tropical and sub-tropical countries are much more sensitive to climate change than temperate agriculture (?). ? conducted a meta-study of papers focused on effects of climate change impacts on crop productivity in Africa and South Asia. The authors concluded that the projected mean change in yield is -8% by the 2050s in both regions. According to a meta-study conducted by ?, losses in global production of wheat, maize and rice are expected with 2 °C global warming. However, with adaptation the simulated yields are on average 7 – 12% higher than without adaptation (?).

A number of important studies have focused on responses to drought related disasters or emergency situations. ? have questioned whether drought is actually a major cause of food insecurity in Ethiopia, Kenya and Somalia. They have analysed causes of the food crisis identified by the humanitarian appeal documents. Majority of these documents have found food availability and food production to be the major causes of food crisis. ? argue that there is a tendency to explain failure of more complex food systems as 'droughts' and insufficient attention is paid to non-climatic drivers such as food prices or conflicts. According to ?, a large share of humanitarian response budget has been focused on emergency food aid in comparison to the share focused on interventions to build resilience. The authors have suggested to increase the budget share focused on agricultural and livestock production to build resilience. Also other authors have promoted shifting from reactive to proactive approach in disaster risk management (???). ? has recommended forecast based financing as it can avoid significant disaster losses. ? has

shown that investments into disaster risk reduction had usually been outweighed by avoided losses. Also ? have argued that risk assessment is essential for disaster risk management.

In the light of these arguments, the goal of the present study is to develop a model which will utilize weather data to assess risk of drought and food security. In addition, we want to find out which particular features of climate or weather are the most important factors affecting the yields and food security. Are the average weather conditions the most important or is it the weather variability or length and number of dry spells during the growing season what matters the most? Answering these and other similar questions should help to improve food security and shift the focus from reactive to proactive approach in drought disaster risk management. We focus on Kenya utilizing an yearly panel dataset of 47 counties over the period of 1981 – 2017.

The main contribution of this study is utilization of the most detailed publicly available weather datasets while accounting for the real life crop calendar and different cropping seasons across the country as practised by Kenyan farmers. Furthermore, we believe that this is the first study which has described application of linear mixed effects models to estimate effects of precipitation and temperature on crop yields in Kenya. Hence, this is the first time that effects of drought and weather on food security have been estimated taking into account the real crop calendar, spatially-heterogeneous nature of Kenya and allowing for the effects of weather to vary across the country.

## 2 Methodology

### 2.1 Measures of food security and drought

Before beginning our research, we had to answer two essential questions: ‘How should we measure food security?’ and ‘How should we measure drought?’ Taking into account the aim of this study, previous literature (see Section 1) and data availability, we were considering three main approaches to measuring food security. In particular, we considered health and utility indicators, food prices and agricultural crop yields as measures of food security. Although some health and utility indicators including the Mid Upper Arm Circumference (MUAC) of children under five, the Coping Strategies Index (see ?) or the NDMA early warning phases are available in the NDMA Early Warning Bulletins, these data are only available since July 2013. The other problem is that the scale of indicators does not seem to be consistent over the counties and time. Furthermore, the values of indicators are often missing and for some months the reports are not present at all.

Using food prices as a measure of food security turned out to be problematic as well. One problem is data availability as the food prices are usually available for market towns but not as county averages. Furthermore, the market data which are available in the publications of ? (KNBS) only cover several years. Another problem is that prices are usually strongly spatially autocorrelated and this could cause the regression estimates to be inconsistent and inefficient unless the model accounts for the spatial autocorrelation structure. Modelling the structure of spatial autocorrelation would require developing and estimation of complicated structural equations and also additional data, which may not be available. Therefore we opted for agricultural crop yields as a measure of food security and

we decided to estimate a single reduced form equation with crop yields as the dependent variable. In particular, we focus on maize yields as this crop is the principal staple food in Kenya and it is grown at 90% of farms in Kenya (?).

The other important question to answer before starting our analysis was: ‘How is drought defined and which definition and measure of drought would be the best for the purpose of this study?’ Various definitions of drought and their role have been reviewed by ? and ?. For an extensive overview of drought indices see ?, ?, ? or ?. ? have quantified, evaluated and compared number of drought indices used for measuring of severity of meteorological, hydrological and agricultural types of drought. For more details about types, measures and indices of drought see Appendix 1.

In the recent period, remote sensing data have been used increasingly to monitor levels of greenness and closely related vegetation conditions (?). Based on these data, various measures (for example the vegetation condition index) have been developed and used for quantifying drought strength and severity (?).

We have considered the Standardised Precipitation Evapotranspiration Index (SPEI) and the vegetation condition index (VCI) as measures of drought in our analysis. However, we did not find the county level SPEI to be as good predictor of food security as other precipitation and temperature measures and computing the SPEI for the entire high resolution weather dataset would be inadequately demanding. We did not use the VCI either as we were not confident enough about the reliability of the available data.<sup>1</sup>

Thus, after conducting a rigorous literature research on measures of drought previously used in the relevant context, we decided to utilize various aggregates of daily precipitation and temperature. The frequency of the maize yield data is yearly while the weather data

---

<sup>1</sup>The values of the VCI which we obtained from the NDMA were not the same as those available by the ? (BOKU). Furthermore, the VCI values did not exactly correspond to the early warning phases as defined by the ? (NDMA).

are daily. Hence, the weather data needed to be aggregated in order to obtain a dataset conformable with the yield data. There are many possible ways of aggregation. Commonly used measures include monthly averages (or monthly totals in case of precipitation) and their variances or standard deviations (???). For example, ? has analysed relationship of crop yields and three measures of precipitation in Sub-Saharan West Africa. These include: (i) Total rainfall during the first month of the period from sowing to harvesting (June) (ii) Total rainfall during the first two months of the period from sowing to harvesting (June and July) and (iii) Total rainfall during the first three months of the period from sowing to harvesting (June, July and August). Based on his results, weather during June and July is the most important for crop yield in Sub-Saharan West Africa (?). Other studies have utilised seasonal totals or means (??) or annual totals or means (?). ? have found sea surface temperature anomalies at various locations and amount of rainfall in July, August and September to be significant for millet crops in Niger. The authors have also considered the maximum air temperature in the hottest month (April) and the minimum air temperature in the coldest month (January) as possible predictors of crops in Niger, but they did not find them significant. Some studies are based on simulated daily extremes, averages, daily measures of variance (???) or yearly extremes (?). Other measures used to explain variability of maize yield include counts of wet and dry days per a defined period, usually a month or a season (????) or length of rainy season (??). Definitions of wet and dry days and rainy season vary. For example, ? have assumed that rainy season begins when the amount of rainfall in three consecutive days reaches at least 25 mm and no dry spell of more than seven days occurs in the following thirty days. According to this study, the end of the rainy season is defined as that rainy day after which rain recorded during 20 days is less than 5 mm.

An important group of studies that have analysed relationship of yield and climate in Sub-Saharan Africa has utilised degree days (????) or number of days with temperature



above certain level or within defined range (??).

Erin Lentz, can we cite her paper?She has not responded to my email

maybe following into data section??

Based on the literature research above and complexity of deriving the measures, we short-listed the following aggregates of the weather data:

- **Precipitation:**

- Seasonal cumulative rainfall
- Seasonal standard deviation
- Seasonal coefficient of variation
- Maximum length of dry spell in number of days
- ...the list to be finished
- ...

- Temperature

- 
- the list to be finished..
- 

## 2.2 Data

- Source of data
- Verbal description of the data, resolution, period, frequency

- Plot(s)/barplots of yield, precip and temp when final data available
- Maybe a table of descriptive statistics or ref to the tale to appendix
- Description how the weather data were aggregated spatially over counties
- Description how the weather data were aggregated temporally - seasons

## 2.3 Linear mixed models

Kenya consists of 47 counties with semi-autonomous county governments (?). As a result of the high degree of county-level autonomy, the policies and regulations often differ across the counties, hence the effects of weather on crop yield are likely to be different across the counties. Therefore, following the standard methodology, we estimated a battery of linear mixed effects models (mixed models) commonly used to analyse longitudinal data (?). Mixed models are suitable for analysis of panel data as they account for the panel structure of the dataset. These types of models include both fixed effects parameters and random effects. Fixed effects are analogous to parameters in a classical linear regression model and value of each effect is assumed to be fixed over all counties (?). On the other hand, random effect are unobserved random variables. There are at least three benefits of treating a set of parameters as a random sample from some distribution. *(i)* Extrapolation of inference to a wider population *(ii)* improved accounting for system uncertainty and *(iii)* efficiency of estimation (??).

Formally, a linear mixed model can be described by the distribution of two vectors of random variables: the response  $\mathcal{Y}$  and the vector of random effects  $\mathcal{B}$ . The distribution of  $\mathcal{B}$  is multivariate normal and the conditional distribution of  $\mathcal{Y}$  given  $\mathcal{B} = \mathbf{b}$  is multivariate normal of a form (??):

$$(\mathcal{Y}|\mathcal{B} = \mathbf{b}) \sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{b}, \sigma^2\mathbf{I}), \quad (1)$$

where  $\mathbf{X}$  is an  $n \times p$  model matrix of fixed effects,  $\beta$  is a  $p$ -dimensional fixed-effects parameter,  $\mathbf{Z}$  is an  $n \times q$  model matrix for the  $q$ -dimensional vector of random-effects variable  $\mathcal{B}$  evaluated at  $\mathbf{b}$  and  $\sigma$  a scale factor. The distribution of  $\mathcal{B}$  can be written as:

$$\mathcal{B} \sim N(0, \Sigma), \quad (2)$$

where  $\Sigma$  is a  $q \times q$  positive semi-definite variance-covariance matrix

### 3 Results and discussion

- Tables of estimates:
  1. Base model which only includes seasonal precipitation and average seasonal temperature
  2. The preferred specification which includes the significant weather aggregates (see the preliminary results in Table 1)
    - For all counties and also separately for Arid or semi-arid (ASAL) and non-ASAL
- Verbal description and interpretation of the results
- Estimates of future impacts
  - Comparison of estimates accuracy of the predictions from the base model and the preferred specification

- \* choose two points in time (past). Imagine we are in the earlier of the two time points and we want to make a prediction of yields in the future point. Compare the predictions of the base model and the predictions of the preferred specification. Here, we assume that we know the forecast of the weather measures- variables which are used in the models. We can use the real weather data instead of the forecast. This will demonstrate why is it important to not only forecast totals/means but also other characteristics (measures) of weather/climate
- \* Graphical comparison, possibly using barplots

## 4 Conclusion

- Besides seasonal totals (averages), there are many weather/climate characteristics (measures) which are important for yield
- Current weather and climate forecasts (in the context of disaster management) are mostly focused on totals in precipitation (and average temperature?)
- There is a strong movement to shift the focus from forecasting merely weather/climate to forecast of weather/climate and its impacts

→ We can stress the importance of focusing the forecast on the climate characteristics (or measures) which we find significant in our models

- Caveats
  - Yield data: Is the collection method consistent over time and space?
  - Livestock

- Cropping calendar is likely to differ across the country

**Table 1: *Mixed effects model: Log of maize yield and weather, ARMA(1,1) errors***

<b>Fixed effects:</b>	<i>All counties</i>		<i>ASAL</i>		<i>non-ASAL</i>	
	$exp(\beta)$	F-value <sup>a</sup>	$exp(\beta)$	F-value <sup>a</sup>	$exp(\beta)$	F-value <sup>a</sup>
Intercept	1.296***	19.916	1.276*	5.230	1.410**	10.061
Prec. total	1.081*	5.402	1.006	0.022	1.278***	19.386
Prec. total sq.	0.973*	4.289	1.004	0.051	0.880***	23.747
Prec. c. of var.	0.924•	3.277	0.969	0.246	0.909	2.231
Dry spell -length	0.935*	4.810	0.833**	6.969	0.988	0.163
Dry spells $\geq 4$ d.	0.939*	4.826	0.855**	8.065	0.989	0.096
Temp. - average	0.819***	12.127	0.808*	5.376	0.878	1.580
Temp. std. dev.	1.043•	3.125	1.039	0.558	1.059 *	5.640
<b>Random effects:</b>						
Intercept						
<i>Number of observations:</i>	1300	698	602			

*Notes:* Standard errors in brackets, 1256 observations

•  $p < 0.1$ ; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

a Marginal (type III) sum of squares. The F-statistics correspond to the sum of squares attributable to each fixed effect.

## Appendix 1 Drought: definitions, measures and indices

According to the international meteorological community, drought can be defined in several ways. In particular, drought is a '*prolonged absence or marked deficiency of precipitation*', a '*deficiency of precipitation that results in water shortage for some activity or for some group*' or a '*period of abnormally dry weather sufficiently prolonged for the lack of precipitation to cause a serious hydrological imbalance*' (??). ? has defined three types of droughts: (i) 'Agricultural drought' which is defined in terms of moisture deficits in upper layer of soil up to about one meter depth (ii) 'meteorological drought' which refers to prolonged deficit of precipitation and (iii) 'hydrological drought' which relates to low streamflow, lake and levels of groundwater. The ? policy statement was later replaced by another statement (?) which besides the three types of drought above, covers also the 'socioeconomic drought' which associates the supply and demand of some economic good with elements of meteorological, agricultural and hydrological drought (??).

? and ? have distinguished two main categories of definitions of drought: (i) conceptual and (ii) operational. Conceptual definitions are dictionary types, usually defining boundaries of the concept of drought<sup>2</sup>. Operational definitions are essential for an effective early warning system. An example of operational definition of agricultural drought can be obtaining the rate of soil water depletion based on precipitation and evapotranspiration rates and expressing these relationships in terms of drought effects on plant behaviour (?).

In order to compare severity of drought across different time periods or geographical locations a numerical measure turns out to be necessary. However, as a result of a large disagreement about a definition of drought, there is no single universal drought index. Instead of that a number of measures of drought has been developed (???).

---

<sup>2</sup>An example of conceptual definition of drought is an 'extended period - a season, a year, or several years of deficient rainfall relative to the statistical multi-year mean for a region' ?.

Examples of early measures of drought are ?, ?, ? or ?. ? suggested to use length of period without 24-h precipitation of 1.27 mm. ? is based on a measure of precipitation over a given time period. ? proposed to measure severity of drought as a length of drought in days where the end of a drought is defined by occurrence of 2.54 mm of precipitation in 48 hours. ? developed the Antecedent Precipitation Index (API) which is based on amount and timing of precipitation and it was used for forecasting of floods. Hence, the API is a reverse drought index.

The study of ? was a significant milestone in the history of quantification of drought severity. ? developed the Palmer Drought Severity Index (PDSI) using a complex water balance model. The PDSI is based on a hydrological accounting system, which incorporate antecedent precipitation, moisture supply and moisture demand (??). As the PDSI suffers from several weaknesses (for details see e.g. ?), other indices were developed in the following decades. These include the standardized precipitation index (SPI) developed by ? and the standardized precipitation evapotranspiration index (SPEI) developed by ?. The SPI specifies observed precipitation as a standardised departure from a chosen probability distribution which models the precipitation data. Values of SPI can be viewed as a multiple of standard deviations by which the observed amount of rainfall deviates from the long-term mean (?).<sup>3</sup> The SPEI is similar to SPI, but unlike SPI, the SPEI includes the role of evapotranspiration (which captures increased temperature). It is based on water balance, therefore it can be compared to the self-calibrated PDSI (?).

## Appendix 3 Tables

---

<sup>3</sup>Can be created for various periods of 1-36 months, usually using monthly data.