Submitted to: *Environ. Res. Lett.*

The Effects of weather on maize yields: New evidence from Kenya

Monika Novackova, Pedram Rowhani, Martin Todd and Dominic Kniveton

Department of Geography, University of Sussex, Falmer, UK

E-mail: monika.novac@gmail.com

November 2018

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

.. to be written later..

1. Introduction

Paragraph 1

- Extreme weather causes disasters → early warning systems have been developed

Paragraph 2

- What weather forecasts (measures) have been used in EWS? *ref. litrature*
  - Mostly seasonal precip. totals and temperature averages

Paragraph 3

- Identify difference between hazard and disaster
  - Not every hazard turns into disaster
  - For a hazard to become a disaster it needs to have **impact**
  - Here, we identify the key metrics which have impact on yield

Paragraph 4

- Crop yield versus climate forecasting

Paragraph 5

- Aim of the paper

2. Methods

...a case study looking at Kenya...

2.1. Data

In this study, we analyzed the relationship between maize yields and climate. Our dataset consists of an yearly panel of 47 counties of Kenya describing the period of $1981 - 2017$. We acquired the county level yield data from the Famine Early Warning

Systems Network (FEWS NET). As for the weather data, we utilized 0.25° resolution precipitation and temperature gridded datasets. The precipitation data come from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) while the temperature data are available at the website of the Berkeley Earth.

Both precipitation and temperature datasets are daily while the yield data are yearly by its nature. Hence, in order to proceed with the analysis, the daily precipitation and temperature data had to be converted into a county-level yearly panel. The first step of this procedure was averaging the gridded weather data over the counties of Kenya resulting in a single daily time series for each county. In the next step, we aggregated the resulting county level time series temporally to obtain yearly time series using the seasonal crop calendar which is available at the website of FEWS NET. We separated the counties of Kenya into two groups, such that the crop calendar can be considered homogeneous within each group. The groups are: *(i)* Arid and semi-arid (ASAL) counties and *(ii)* non-ASAL counties as defined by the ASAL Stakeholders Forum (ASF). For the list of counties in each of the groups see table 1 in the appendix. ‡.

In terms of weather, the most relevant periods for yield are the rainy seasons. Therefore, we only considered weather during the rainy seasons for our analysis. According to FEWS NET, the ASAL counties have two distinct rainy seasons while the non-ASAL counties have one longer rainy seasons. The two rainy seasons of the ASAL counties are: *(i)* March-May and *(ii)* October-December. In the non-ASAL area, the rainy season is in March-August.

Hence, for the ASAL counties, we aggregated daily precipitation and temperature

‡ The crop calendar is probably not exactly the same for all counties within each group, however, it is similar enough so that it can be considered homogeneous within the groups. To a certain extent, the crop calendar is likely to differ even within a single county. However, considering the information provided by ASF, FEWS NET and accounting for our intermediate results, we opted for the two groups as a compromise between accuracy and complexity.

over October, November and December of the previous year and May, April, May of the current year to get single value of the weather variables for each county and year. For the non-ASAL counties we aggregated the daily weather values over March, April, May, June, July and August to obtain yearly values.

2.2. Statistical approach

- *We used commonly used measures of weather/drought (Only mention the significant weather measures/variables)*
- Describe the temporal aggregation of the weather variables, seasons

Kenya consists of 47 counties with semi-autonomous county governments (Barasa, Manyara, Molyneux & Tsofa 2017). 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 (also known as mixed models) commonly used to analyse longitudinal data (Bates, Pinheiro, Pinheiro & Bates 2000). 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 (Bates 2010). On the other hand, random effects 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 $\mathscr{Y}$ and the vector of random effects $\mathscr{B}$. The distribution of $\mathscr{B}$ is multivariate normal and the conditional distribution of $\mathscr{Y}$ given $\mathscr{B} = \mathbf{b}$ is multivariate normal of a form

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

where $\mathbf{X}$ is an $n \times p$ model matrix of fixed effects, β 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 $\mathscr{B}$ evaluated at $\mathbf{b}$ and σ a scale factor. The distribution of $\mathscr{B}$ can be written as:

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

where $\mathbf{\Sigma}$ is a $q \times q$ positive semi-definite variance-covariance matrix.

- Add a description of residual and other diagnostic tests (AIC, VIF, autocorrelation)
- Describe the procedure that I have applied to find the preferred combination of fixed effects and random effects
- Describe the procedure that I have applied to find the preferred way of modelling the correlation structure in errors (ARMA errors)

3. Results

- Tables of estimates of the preferred specification which includes the significant weather variables (see the preliminary results in Table 1)
  - For all counties
  - For Arid and semi-araid (ASAL) counties
  - For non-ASAL counties

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

	All counties		***ASAL***		***non-ASAL***	
Fixed effects:	**Estimate**	**F-value[a]**	**Estimate**	**F-value[a]**	**Estimate**	**F-value[a]**
Intercept	0.259***	19.916	0.243*	5.230	0.344**	10.061
Prec. total	0.078*	5.402	0.006	0.022	0.246***	19.386
Prec. total sq.	−0.028*	4.289	0.004	0.051	−0.128***	23.747
Prec. c. of var.	−0.079•	3.277	−0.031	0.246	−0.095	2.231
Dry spell -length	−0.067*	4.810	−0.183**	6.969	−0.012	0.163
Dry spells $\geq$ 4 d.	−0.063*	4.826	−0.157**	8.065	−0.011	0.096
Temp. - average	−0.199***	12.127	−0.213*	5.376	−0.130	1.580
Temp. std. dev.	0.042•	3.125	0.038	0.558	0.057 *	5.640
Random effects:						
Intercept						
Number of observations:	1300		698		602	

Notes: Standard errors in brackets; • $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

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

- Verbal description and interpretation of the results. Discussing goodness of fit using various criteria such as AIC or alternatives to R^2.
- If we get the yield data for the period from 2015 onwards: Out of sample predictions and comparison with the real data

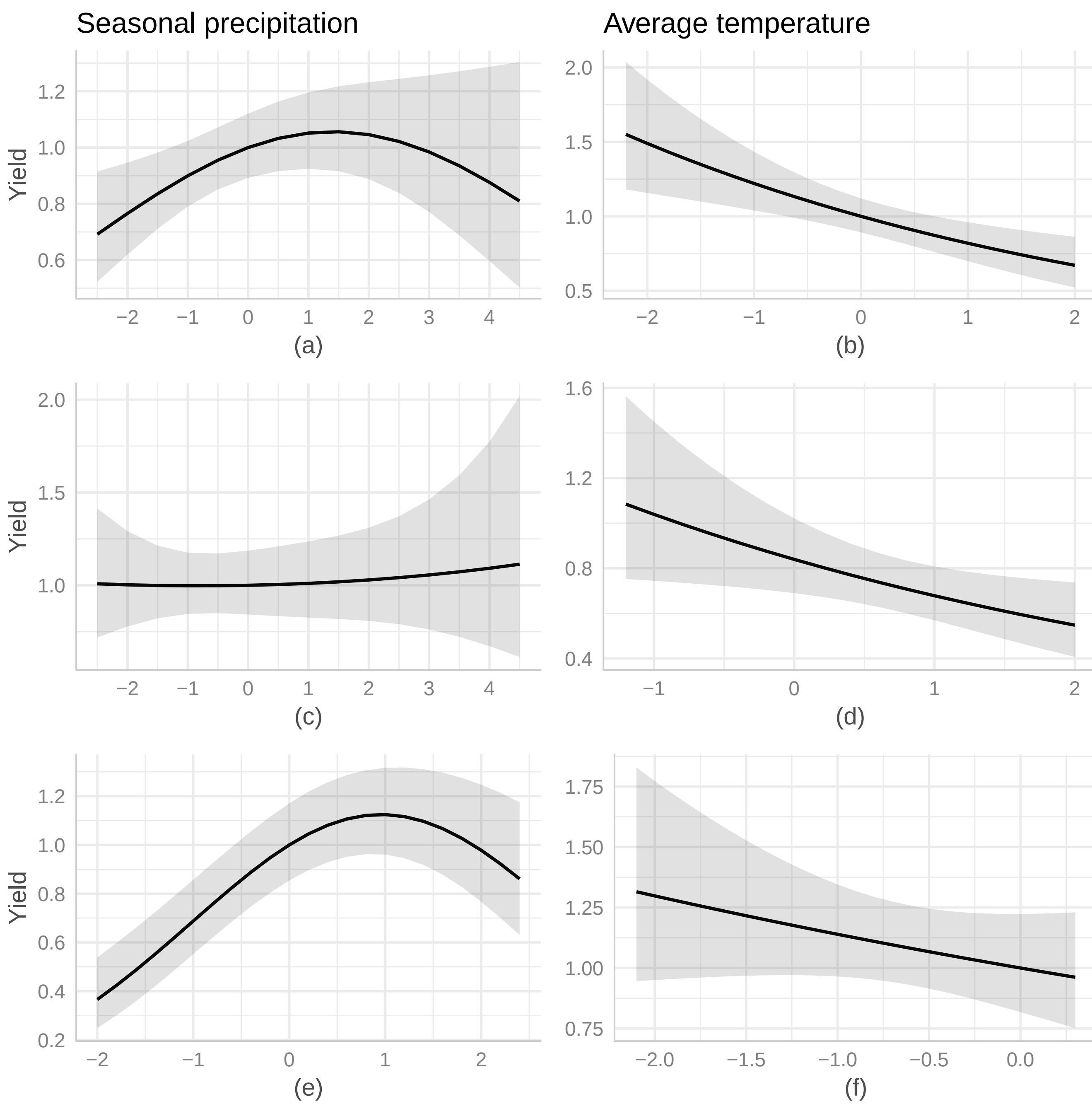

Figure 1. Predicted multiplicative marginal effects of seasonal precipitation (left column) and average seasonal temperature (right column) on maize yields. The first row (a, b) represents the model for all counties, the second row (c, d) is based on the subsample of arid and semi-arid counties (ASAL) and the third row (e, f) represents the model for the non-ASAL counties. Precipitation and temperature (x-axis) are in multiples of their standard deviations. The effects are multiplicative as the models are in log-linear form.

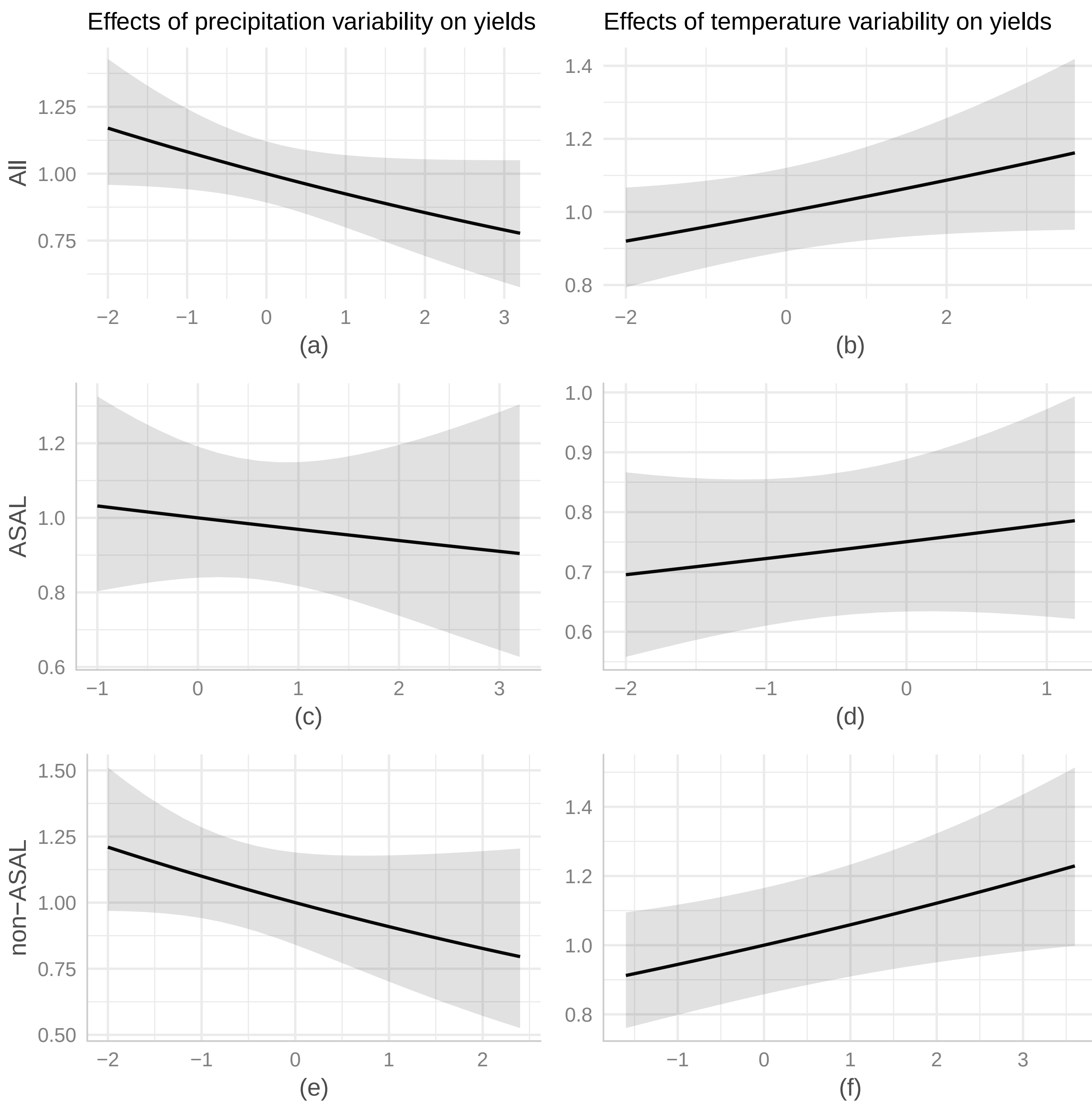

Figure 2. Predicted multiplicative marginal effects of coefficient of variation (CV) of precipitation (left column) and standard deviation (SD) of temperature (right column) on maize yields. The first row (a, b) represents the model for all counties, the second row (c, d) is based on the subsample of arid and semi-arid counties (ASAL) and the third row (e, f) represents the model for the non-ASAL counties. CV of precipitation and SD of temperature (x-axis) are in multiples of their standard deviations. The effects are multiplicative as the models are in log-linear form.

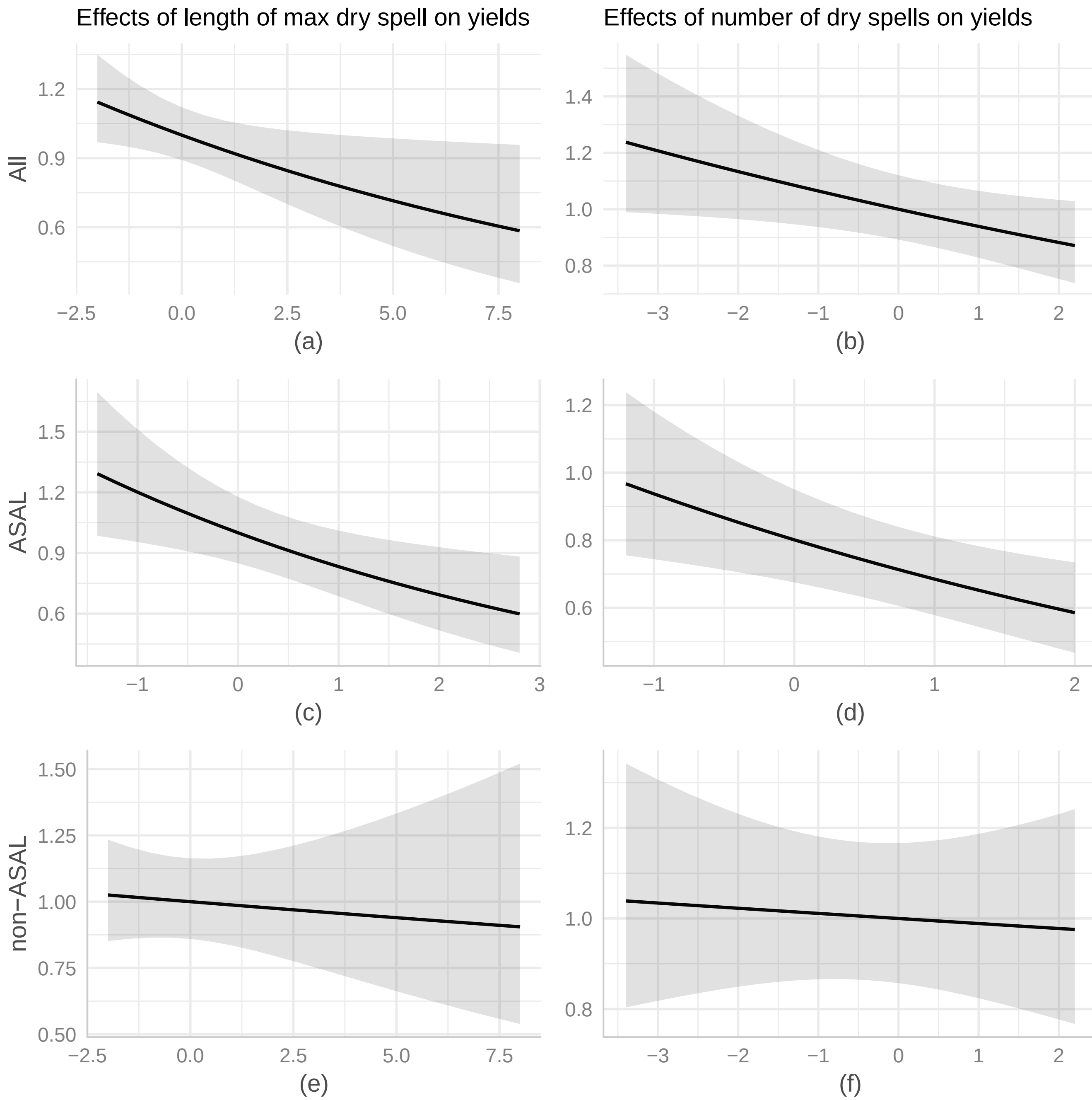

Figure 3. Predicted multiplicative marginal effects of length of maximum dry spell in days (left column) and number of dry spells lasting for four days or more (right column) on maize yields. The first row (a, b) represents the model for all counties, the second row (c, d) is based on the subsample of arid and semi-arid counties (ASAL) and the third row (e, f) represents the model for the non-ASAL counties. Maximum length of dry spell and number of dry spells (x-axis) are in multiples of their standard deviations. The effects are multiplicative as the models are in log-linear form.

Appendix

should be at the end of the main text but before list of references

Table 1. List of Arid and semi-arid (ASAL) and non-ASAL counties

ASAL:	Baringo, Embu, Garissa, Isiolo, Kajiado, Kilifi Kitui, Kwale, Laikipia, Lamu, Makueni, Mandera, Marsabit, Meru, Mombasa, Narok, Nyeri, Samburu, Taita-Taveta, Tana River, Tharaka Nithi, Turkana, Wajir, West Pokot
non-ASAL:	Bomet, Bungoma, Busia, Elgeyo Marakwet, Homa Bay, Kakamega, Kericho Kiambu, Kirinyaga, Kisii, Kisumu, Machakos, Migori, Murang'a, Nakuru, Nyamira, Nyandarua, Siaya, Trans Nzoia, Uasin Gishu, Vihiga

Possibly include a table of all values which I get from the lme or lme4 models summary in R, that is. correlation of the fixed effects etcetera

Maybe also a table with standard errors here

Table 2. ***Mixed effects model:***
Log of maize yield and weather, ARMA(1,1) errors

	All counties		***ASAL***		***non-ASAL***	
Fixed effects:	$exp(\beta)$	F-value[a]	$exp(\beta)$	F-value[a]	$exp(\beta)$	F-value[a]
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
Random effects:						
Intercept						
Number of observations:	1300		698		602	

Notes: Standard errors in brackets; • $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

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

References

Barasa, E., Manyara, A., Molyneux, S. & Tsofa, B. (2017).

Bates, D. M. (2010). lme4: Mixed-effects modeling with R.

Bates, J., Pinheiro, J., Pinheiro, J. & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing, Springer New York.

Online Source: (Accessed in December 2018). Asal Stakeholders Forum (ASF).
URL: *http://www.asalforum.or.ke/*

Data Source: Berkeley Earth. Accessed in October 2018.
URL: *http://berkeleyearth.org/data/*

Data Source: Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS). Accessed in October 2018.
URL: *http://chg.geog.ucsb.edu/data/chirps/*

Data Source: Famine Early Warning Systems Network (FEWS NET). Accessed in October 2018.
URL: *http://fews.net/*