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...

Our dataset consists of an yearly panel of 47 counties of Kenya describing the period of 1981-2017.

2.1. Data

- Source of the climate data: BOKU and Berkeley Earth

- Source of the yield data (Kenya MoA)

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

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

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:

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

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:	*exp*(β)	**F-value**[a]	*exp*(β)	**F-value**[a]	*exp*(β)	**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$^{\bullet}$	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$^{\bullet}$	3.125	1.039	0.558	1.059 $*$	5.640
Random effects:						
Intercept						
Number of observations:	1300		698		602	

Notes: Standard errors in brackets; $^{\bullet}$ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

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

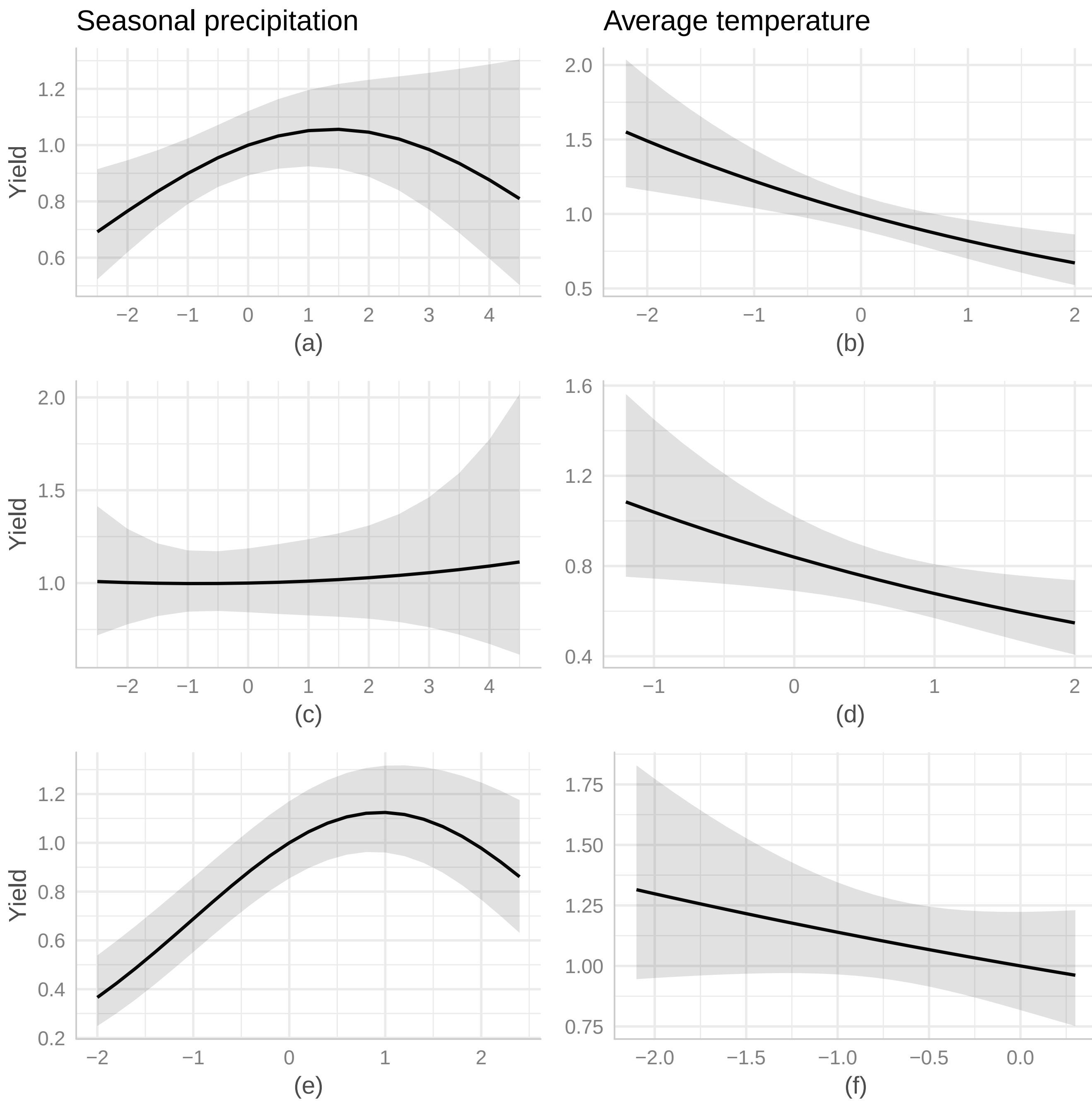

Figure 1. Predicted multiplicative marginal effects of seasonal precipitation (left column) and average seasonal temperature (right column) on maize yields. The first row (a, b) represents the model for all counties, the second row (c, d) is based on the subsample of arid and semi-arid counties (ASAL) 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.

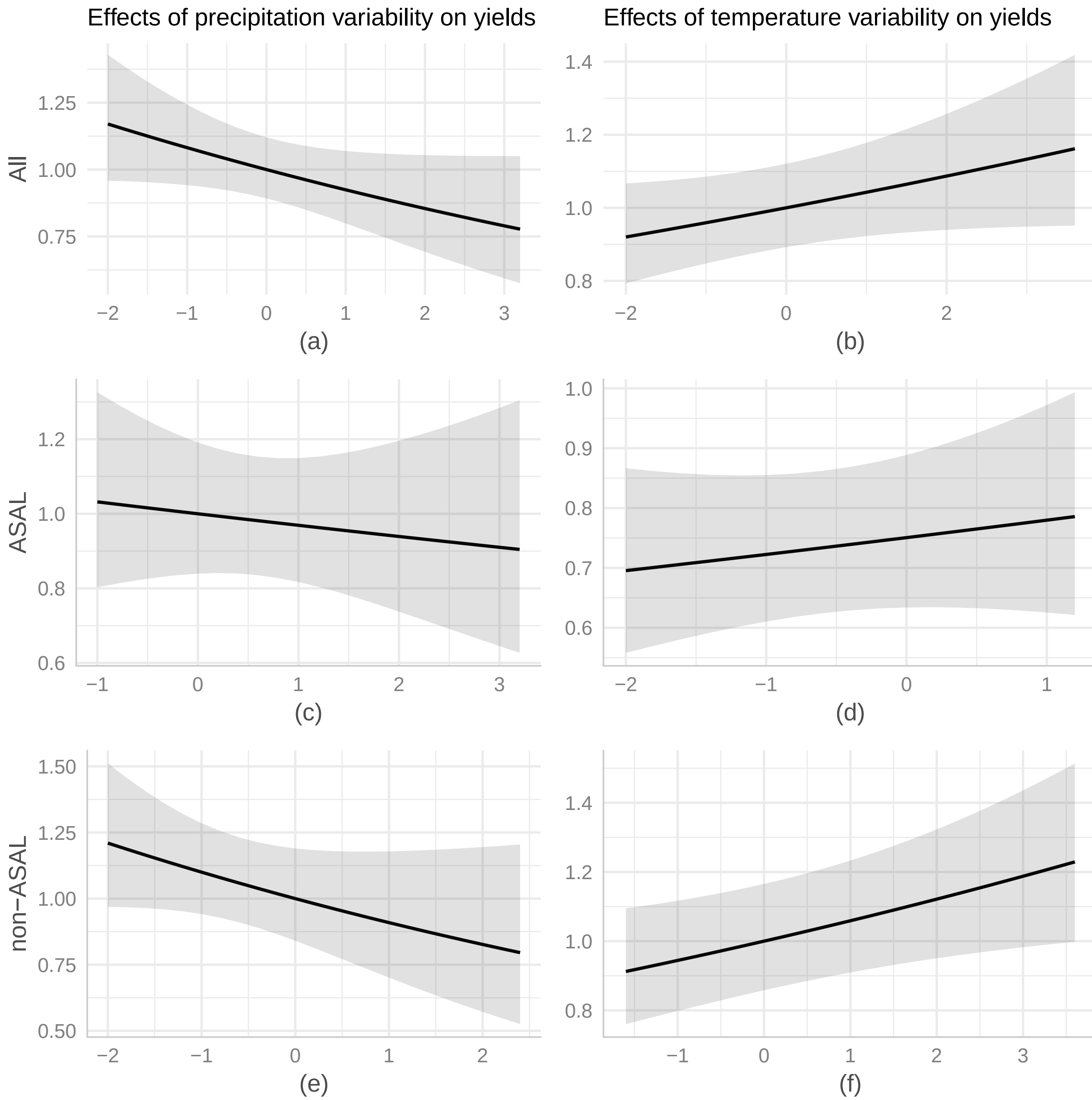

Figure 2. Predicted multiplicative marginal effects of coefficient of variation (CV) of precipitation (left column) and standard deviation (SD) of temperature (right column) on maize yields. The first row (a, b) represents the model for all counties, the second row (c, d) is based on the subsample of 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.

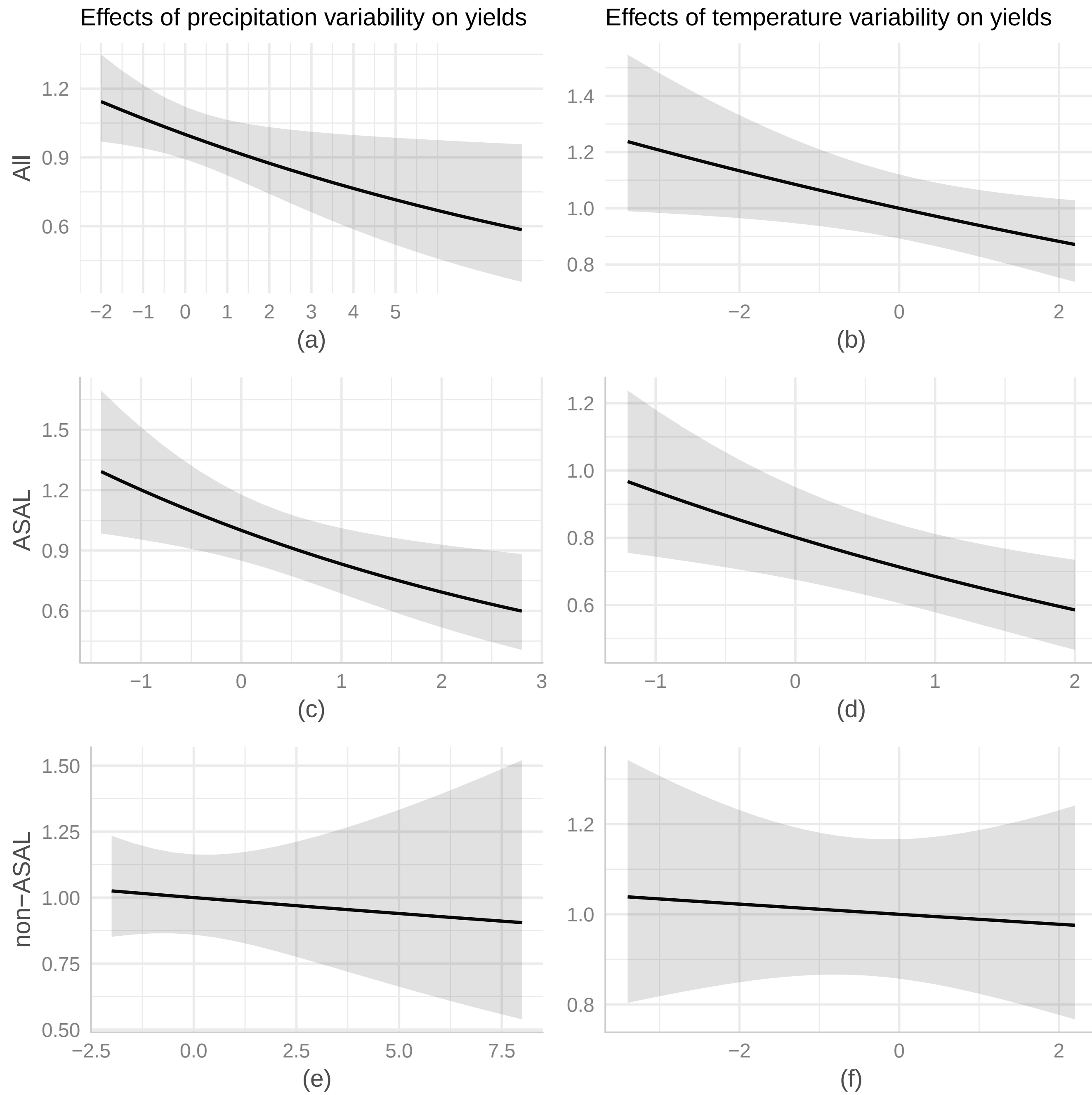

Figure 3. Predicted multiplicative marginal effects of coefficient of variation (CV) of precipitation (left column) and standard deviation (SD) of temperature (right column) on maize yields. The first row (a, b) represents the model for all counties, the second row (c, d) is based on the subsample of 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.

Appendix

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

BOKU *Data Source:* University of Natural Resources and Life Sciences Vienna. Accessed in October 2018.

URL: *http://ivfl-info.boku.ac.at/index.php/homepage?view=featured*

Earth, B. *Data Source:* Berkeley Earth. Accessed in October 2018.

URL: *http://berkeleyearth.org/data/*