Submitted to: *Environ. Res. Lett.*

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

Monika Novackova, Martin Todd, Dominic Kniveton and Pedram Rowhani

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 yearly yield data from the Famine Early

Warning Systems Network (FEWS NET). Regarding the weather data, we exploited 0.25° resolution precipitation and temperature gridded datasets. The precipitation data were obtained from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) while the temperature data are available at the website of the Berkeley Earth. We calculated spatial averages of the gridded data over the counties to get a single value for each county and each point in time. The frequency of precipitation and temperature data was daily, therefore, further aggregation was needed to obtain yearly values corresponding to the yearly frequency of the yield data. Hence, each weather characteristic was further aggregated over years resulting in a county-level panel conformable with the yield data.

There are two predominant rainfall regimes in Kenya. The arid and semi-arid (ASAL) counties exhibit mostly bi-modal precipitation patterns with long rains lasting from March to May and short rains occurring between October and December. In the non-ASAL counties, the single rainy season usually starts in March and lasts until August. Following the precipitation patterns and closely related planting and harvesting calendar, we computed yearly values of various weather characteristics as follows: For the ASAL counties, we used daily data covering October, November and December of the previous year and March, April, May of the current year and for the non-ASAL counties we used daily data covering March to August of the current year.

Following the procedure described above, we calculated a number of previously used characteristics of precipitation and temperature including indicators of floods. The measures which we found significant were: Total seasonal precipitation and its squared values, average seasonal temperature, coefficient of variation of seasonal precipitation, coefficient of variation of average seasonal temperature (converted into Kelvin because other scales of temperature such us Celsius are interval and coefficient of variation does not have any meaning for data measured on an interval scale), maximum length of dry spell in number of days and the number of dry spells lasting for four days or more (we considered every day when precipitation didn't exceed 1 mm as a dry day). We did not include squared values of the average seasonal temperature as it was not significant in our models and there is not much evidence of non-linear relationship between yield and temperature in the past literature.

2.2. Statistical approach

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.

we checked this approach by drop1 =anova testing and it turns out that our model is better than each model without one of the variables..Also the automatic step procedure, which we run for comparison gave us almost the same model (forcing the seasonal precip and average temperature). The only varibale which was extra was MaxT and we decided to remove it because of high VIF - or correlation with average temp...Also mention add1

According to a histogram, the shape of our maize yield data was closer to a log-normal than to a normal distribution. Therefore, following the common practice, the maize yield values were log-transformed before conducting the analysis. Furthermore, we scaled all predictors by subtracting mean and dividing them by standard deviation to avoid convergence problems.

According to the conditional Lagrange multiplier (LM) test developed by Baltagi & Li (1991) and Baltagi & Li (1995), the errors exhibited a within group autcorrelation structure in our models (the p-value of the LM test was 1.1×10^{-13}). To further investigate the autocorrelation structure of the errors, we estimated a number of models (with the subset of predictors chosen as described above) with ARMA(p,q) error structure. In particular, we estimated all variants of our model with ARMA(p,q) errors such that $p \leq 2$ and $q \leq 2$. Comparing the AIC criteria and using the corresponding likelihood ratio statistics, we found that the most appropriate error correlation structure was ARMA(1,1). The value of the AIC criterion of the model with ARMA(1,1) error structure was 2122.2 while the AIC was 2129.2 for the model with ARMA(1,0) errors. The p-value of the likelihood ratio test of the comparison of the model with ARMA(1,1) errors and the model with ARMA(1,0) errors was smaller than 1×10^{-4}, hence the ARMA(1,1) error structure turned out to be a better fit. The AIC criteria and the likelihood ratio tests of all the models with ARMA(p,q) errors are summarised in table 2 in the appendix.

To find the most suitable set of fixed effects, we adapted a stepwise selection procedure to choose a model with the smallest value of the Akaikes Information Criterion (AIC). We started with a model which included the complete set of our weather measures in the fixed effects and random intercepts and we performed a search through the subsets off the fixed effects minimizing the AIC and allowing for steps in both 'backward' and 'forward' directions (removing and adding predictors). We decided to not to include the number of heatwave days and average maximum daily temperature although they slightly improve the AIC if present. The reason for not including them is that they are strongly correlated with average seasonal temperature (the value of correlation coefficient is 0.981 for the maximum daily temperature and 0.517 for the number of heatwave days and the p-values of both of them are smaller than 1×10^{-6}) and they are not significant at the level of significance $\alpha = 0.05$ (their p-values were 0.121 for the maximum daily temperature and 0.075 for the number of heatwave days). Furthermore, including them only improved AIC by 4 (when estimated by maximum likelihood as required for a meaningful comparison) and the signs of these variables were positive in contrary to our assumption. Hence, we concluded that the decrease in AIC and positive signs are likely to be result of their correlation with average temperature or other omitted variable and we did not include them in our model.

We further applied a series of likelihood ratio tests to test for significance of random slopes, however no subset of random slopes improved the fit, hence, our final model only included intercepts in random effects.

To verify our results, we adapted an alternative step-down model building approach using the Satterthwaites method to determine the p-values of the individual t-tests (Kuznetsova, Brockhoff & Christensen 2017). To be more specific, we started with a model which included the complete set of our weather measures in both fixed effects and random effects. In the first part of this procedure the insignificant random effects were removed one by one until only significant predictors remained. After this, the insignificant fixed effects were removed one by one until only significant predictors remained. For random effects we considered the level of significance $\alpha = 0.1$ and for fixed effects we opted for $\alpha = 0.05$ (Kuznetsova et al. 2017). Applying this method we chose the same model as according to the AIC-based stepwise method described above.

3. Results

Besides our main model, we estimated the same specification for two additional subsamples of Kenyan counties. In particular, we estimated the model for the subsample of ASAL counties and for the subsample of non-ASAL counties separately to compare the effects of various weather characteristics across the two areas with significantly different climate. The estimates of all three models are summarised in table 1. The first two columns represent the model for all counties, the third and fourth columns include estimates for the subsample of ASAL counties and the last two columns describe the estimates based on the subsample of non-ASAL counties. All predictors were standardised, therefore the estimates and the corresponding p-values can be interpreted as measures of relative importance.

As an additional measure of relative importance we included F-values of the type *III* analysis of variance (ANOVA) in table 1. The F-tests were conducted using the marginal rather than sequential sum of squares. That is, for each variable, the F-test is a test of significance of that part of explained sum of squares which can be contributed to the particular explanatory variable. In other worlds, the F-value is a test statistic of comparison of the preferred model (with all the variables in table 1) and the model without the explanatory variable in question but including all other predictors. The type *III* ANOVA does not depend on the order in which the effects are entered in the model.

The preferred specification has a log-linear functional form, therefore the estimates in table 1 are not directly interpretable as marginal effects and the effects on yield are multiplicative. The exponents of the estimates (which can be interpreted as additive marginal effects on yield) can be found in table 3 in the appendix.

First, we will focus on the estimates based on the sample of all counties, that is the first two columns in table 1. Consistently with our assumptions, seasonal precipitation has mostly positive and significant effect on maize yield while average temperature has negative effects on yields.

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.	$-0.079^{\bullet}$	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.	$0.042^{\bullet}$	3.125	0.038	0.558	$0.057\ *$	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.

We can see in table 1 that the squared seasonal precipitation is negative while the linear precipitation is posititve,

affects the

Besides seasonal precipitation and average temperature, the length and number of dry spells

- Verbal description and interpretation of the results. Discussing goodness of fit using various criteria such as AIC or alternatives to R^2.
- relative importance of the individual variables and its measures

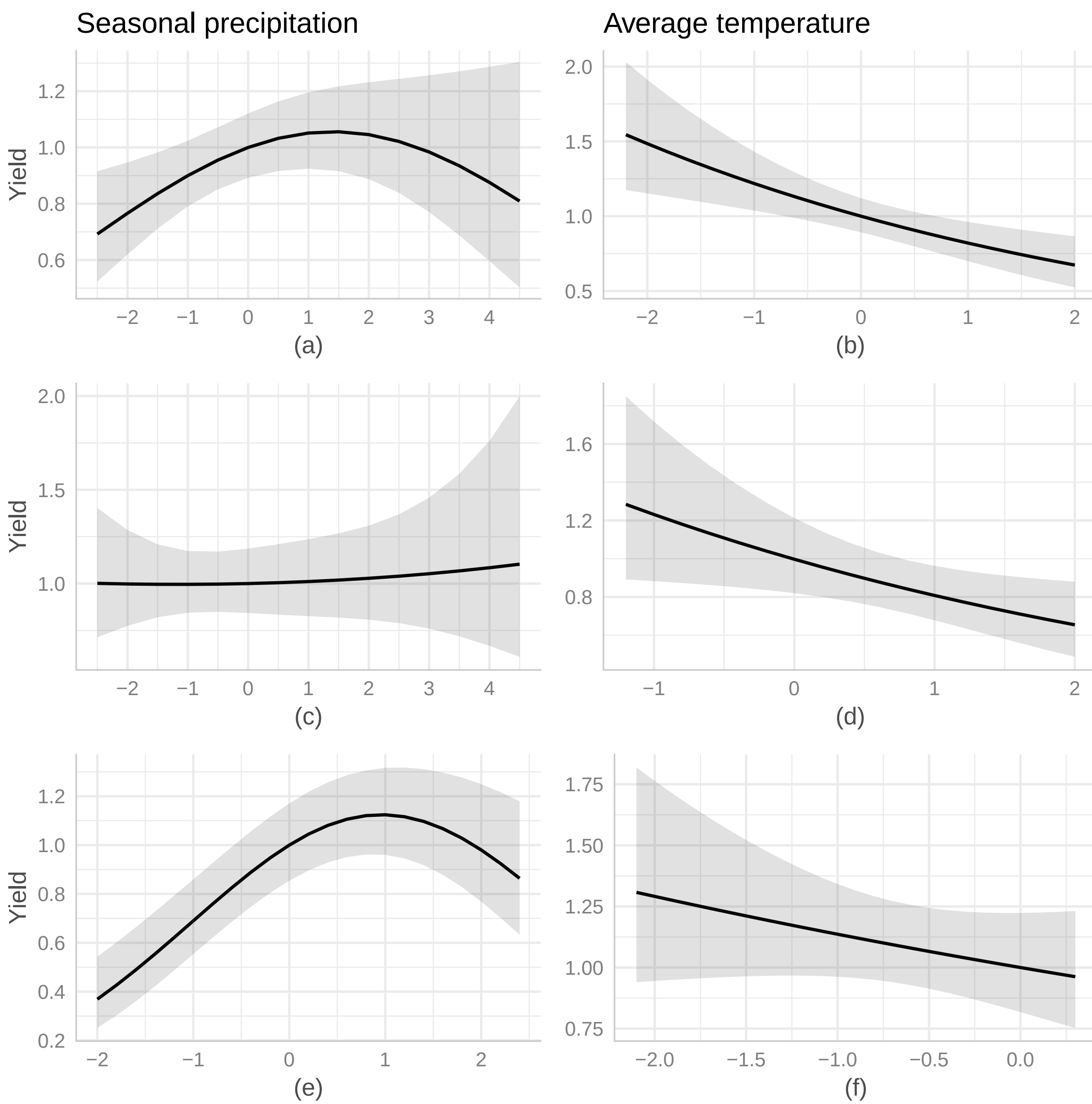

Figure 1. Predicted multiplicative marginal effects of seasonal precipitation (left column) and average seasonal temperature (right column) on maize yields. The first row (a, b) represents the model for all counties, the second row (c, d) is based on the subsample of arid and semi-arid counties (ASAL)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

- show VIF
- If we get the yield data for the period from 2015 onwards: Out of sample predictions and comparison with the real data

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

Table 2. Comparison of models with different error autcorrelation structure

		Likelihood ratio vs. ARMA(1,1)[a]	
Error autocorrelation structure	**AIC**	**Statistic**	**p-value**
None	2205.4	87.17	$< 1 \times 10^{-4}$
ARMA(1,0)	2129.2	8.99	0.0027
ARMA(0,1)	2145.0	24.73	$< 1 \times 10^{-4}$
ARMA(1,1)	2122.2	− − −	− − −
ARMA(2,1)	2124.1	0.15	0.6990
ARMA(1,2)	2124.2	0.01	0.9181
ARMA(2,2)	2125.9	0.31	0.8549

[a] Likelihood ratio test of a comparison of the model in each row against the ARMA(1,1) error structure model. ARMA(1,1) error structure seems to be the most suitable as all lower-order correlation structure models are rejected against ARMA(1,1) while ARMA(1,1) is not rejected against any of the higher order structures.

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 3. ***Mixed effects model:*** *exponents of the coefficient estimates*
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.

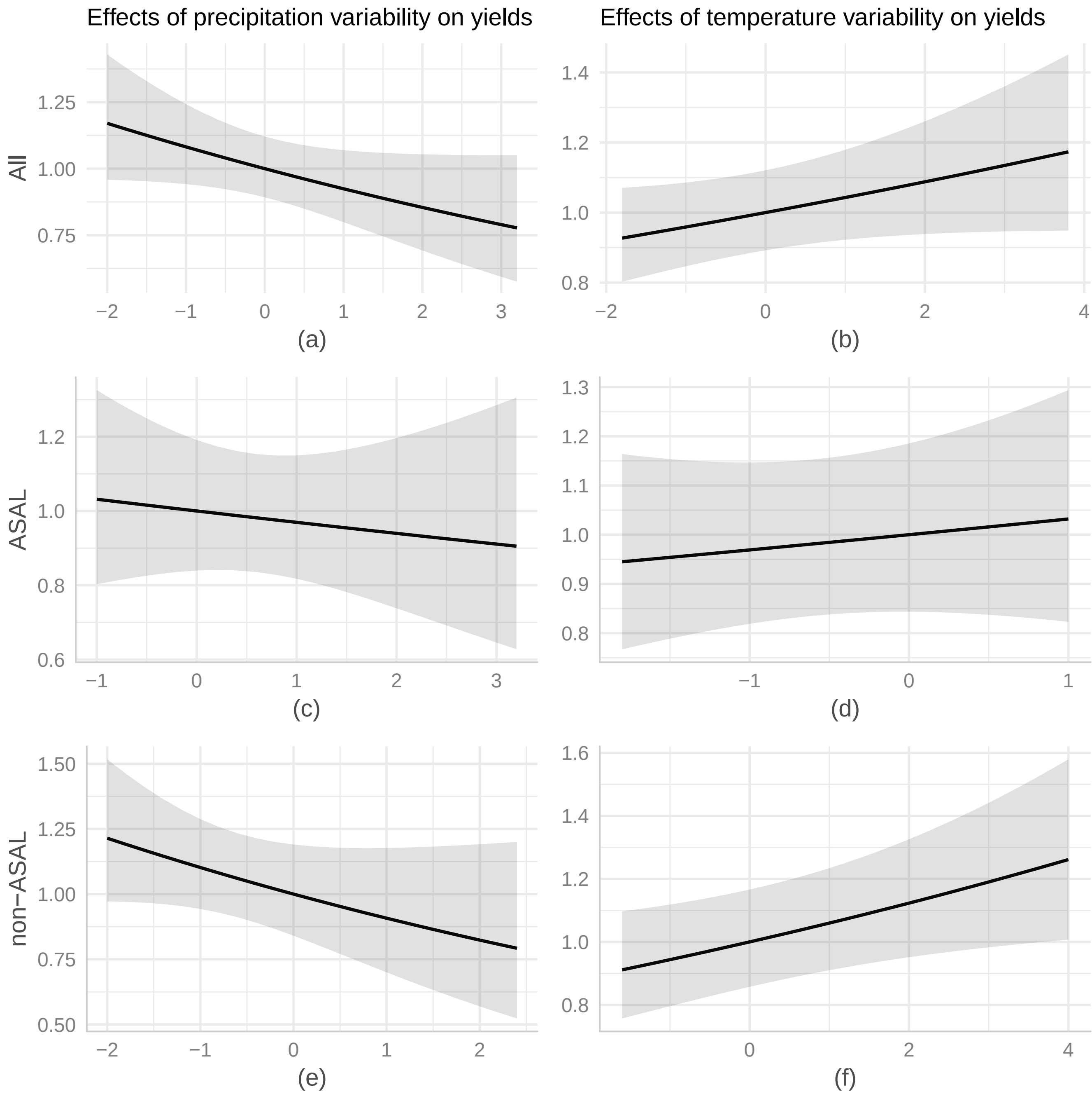

Figure 1. Predicted multiplicative marginal effects of coefficient of variation (CV) of precipitation (left column) and standard deviation (SD) of temperature (right column) on maize yields. The first row (a, b) represents the model for all counties, the second row (c, d) is based on the subsample of 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.

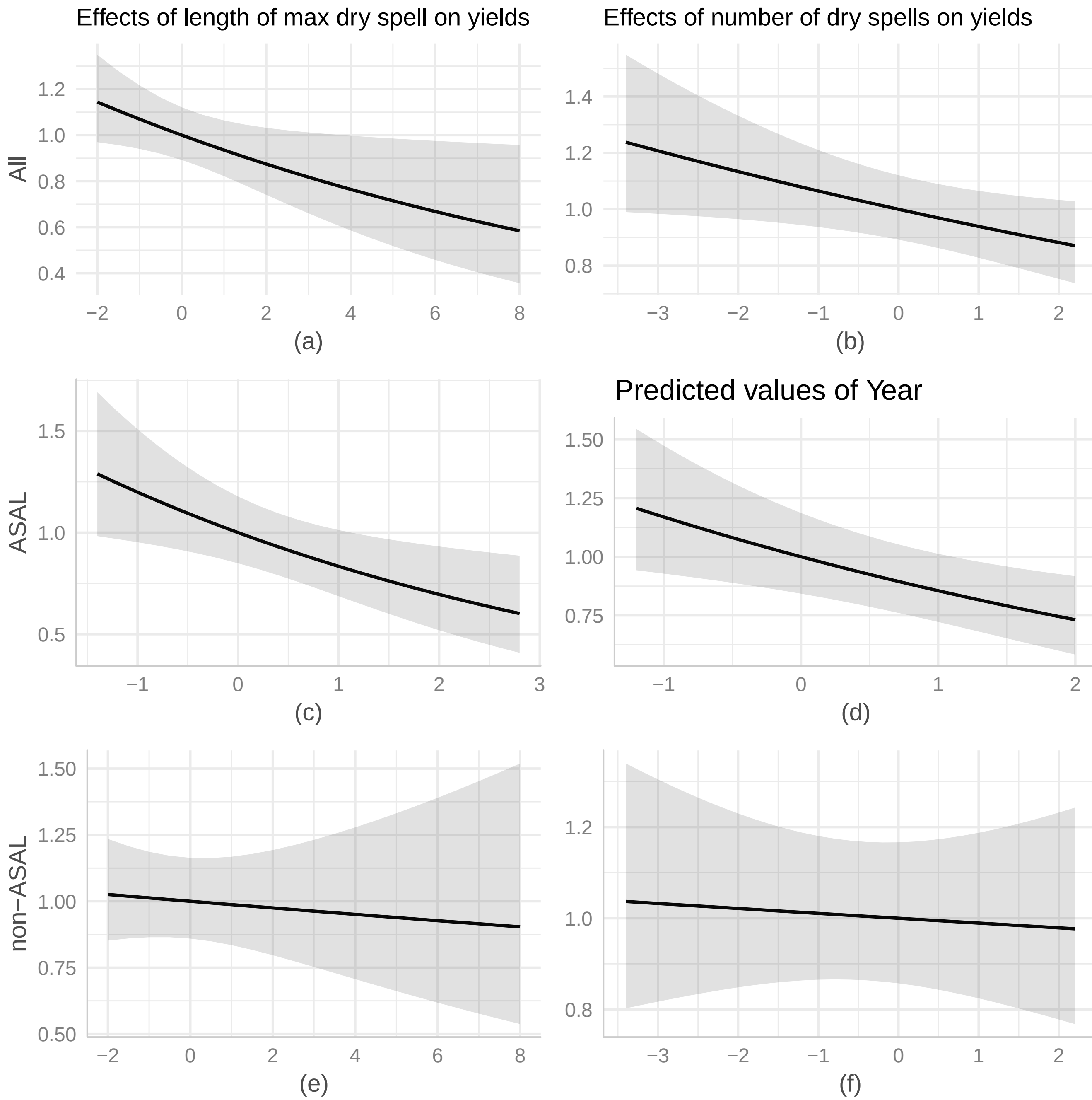

Figure 2. Predicted multiplicative marginal effects of length of maximum dry spell in days (left column) and number of dry spells lasting for four days or more (right column) on maize yields. The first row (a, b) represents the model for all counties, the second row (c, d) is based on the subsample of 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.

References

Baltagi, B. H. & Li, Q. (1991). A joint test for serial correlation and random individual effects, *Statistics & Probability Letters* **11**(3): 277 – 280.

Baltagi, B. H. & Li, Q. (1995). Testing ar(1) against ma(1) disturbances in an error component model, *Journal of Econometrics* **68**(1): 133 – 151.

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.

Kuznetsova, A., Brockhoff, P. & Christensen, R. (2017). lmertest package: Tests in linear mixed effects models, *Journal of Statistical Software, Articles* **82**(13).

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/*