Modelling the relationship between maize yields and precipitation and temperature

Time period: 1999- 2014

- At the moment, the aggregated climate/weather data which I was using are for the period 1999-2018. The yields data are from 1970-2014

Weather data:

1. Precipitation:
   - Z-score at the location where precipitation corresponds to the 10th percentile in each county (I opted for z-scores rather than the raw data because the maximum likelihood algorithm often fails to converge using the raw data.)
   - Monthly frequency
2. Temperature:
   - Z-score at the location where temperature corresponds to the 90th percentile in each county
   - Monthly frequency

Maize yield data:

- MT (= metric tons) per hectare
- Yearly frequency
- Obtained from Gideon Galu from the FEWS NET. The original source of the most of the data is the Ministry of Agriculture.

Precipitation and temperature data aggregation:

- The frequency of the yields data is yearly while the weather data are monthly. We need the frequency of both datasets to be equal for estimating the panel models.
  -> temperature and precipitation need to be aggregated to obtain yearly values
- For each county and year, we aggregate the weather data over the months of planting and growing based no the following seasonal calendar: http://fews.net/east-africa/kenya/seasonal-calendar/december-2013

- According to the seasonal calendar, the counties can be divided into two groups: Eastern and Western.

  - **Eastern counties:**
    - Provinces: Eastern, North Eastern, Coast

        - Two planting and harvesting seasons
        - The yearly climate measures obtained as averages over **November and December of the previous year and January, February, March, April, May, June, July, August and September of the current year** (the months of planting and growing seasons).

    - **Western counties:**
        - Provinces: Rift Valley, Western, Nyanza, Central
        - One harvesting and planting season
        - Yearly climate measures obtained as averages over **May, June, July, August and September** (the months of planting and growing seasons.)

- Besides the average climate measures above, the **coefficients of variation** for both climate and temperature were included in the first models.
    - The coefficients of variation were calculated over the same months as described above for the means (based on the seasonal calendar: http://fews.net/east-africa/kenya/seasonal-calendar/december-2013)
    - The coefficients of variation turn out to be insignificant

First results: Selected specifications of the mixed-effects models

1. No weights

The best specifications of the error structure based on LR tests of serial correlation, and LR tests of random effects:

a) AR(1) errors

lme(Yield~1+PrecZscore +TempZscore , random= ~ PrecZscore + TempZscore|ID, correlation = corAR1(0, form= ~ Year|ID))

$$Y_{it} = 1.685 + \alpha_i + 0.155P_{it} + B_i^{Prec}P_{it} - 0.137T_{it} + B_i^{Temp}T_{it} + u_{it}$$

$$u_{it} = 0.354u_{it-1} + \epsilon_{it} \qquad \epsilon_{it} \sim \mathrm{IID}(0, \sigma_\epsilon^2)$$

b) MA(3) errors

lme(Yield~1+PrecZscore +TempZscore , random= ~ PrecZscore + TempZscore|ID, correlation = corARMA(0, form = ~ as.numeric(Year)|ID, p=0,q=3))

$$Y_{it} = 1.692 + \alpha_i + 0.162P_{it} + \mathcal{B}_i^{Prec}P_{it} - 0.141T_{it} + \mathcal{B}_i^{Temp}T_{it} + u_{it}$$

$$u_{it} = \epsilon_{it} + 0.293\epsilon_{it-1} + 0.198\epsilon_{it-2} + 0.168\epsilon_{it-3} \qquad \epsilon_{it} \sim \text{IID}(0, \sigma_\epsilon^2)$$

No weights

Error structure

R function *pvcm* from the *plm* package (no AR or MA)

Fixed effects	**No AR**	**AR(1)**	**MA(3)**	**Swamy (1970)**	**Counties separately (mean)**
Intercept	1.647***	1.685***	1.692***	1.610***	1.636
Precipitation (z – score)	0.171***	0.155***	0.162***	0.196***	0.201
Temperature (z – score)	-0.109*	-0.137**	-0.141**	-0.089 (p-val= 0.141)	-0.106
AIC	1180.105	1130.211	1122.896		
BIC	1225.900	1180.522	1182.354		

ANOVA: Precipitation explains much bigger part of the variation than temperature (the values of the F-statistics: 44.4 and 7.8).

2. Weights: Area of cropland

```
lme(Yield~1+PrecZscore +cv_Prec, random= ~PrecZscore+TempZscore|ID ,
weights=~Area, correlation= corARMA(form = ~ Year|ID, p=2,q=2))
```

The best specification of the error structure based on LR tests of serial correlation, and LR tests of random effects:

ARMA(2,2) errors:

$$Y_{it} = 1.036 + \alpha_i + 0.187P_{it} + \mathcal{B}_i^{Prec}P_{it} + 0.143vcP_{it} + \mathcal{B}_i^{Temp}T_{it} + u_{it}$$

$$u_{it} = 0.663u_{it-1} + 0.332u_{it-2} - 0.367\epsilon_{it-1} - 0.533\epsilon_{it-2} + \epsilon_{it}$$
$$\epsilon_{it} \sim \mathrm{IID}(0, \sigma_\epsilon^2)$$

AIC: 1817.509 **BIC:** 1881.509

Weights = Area

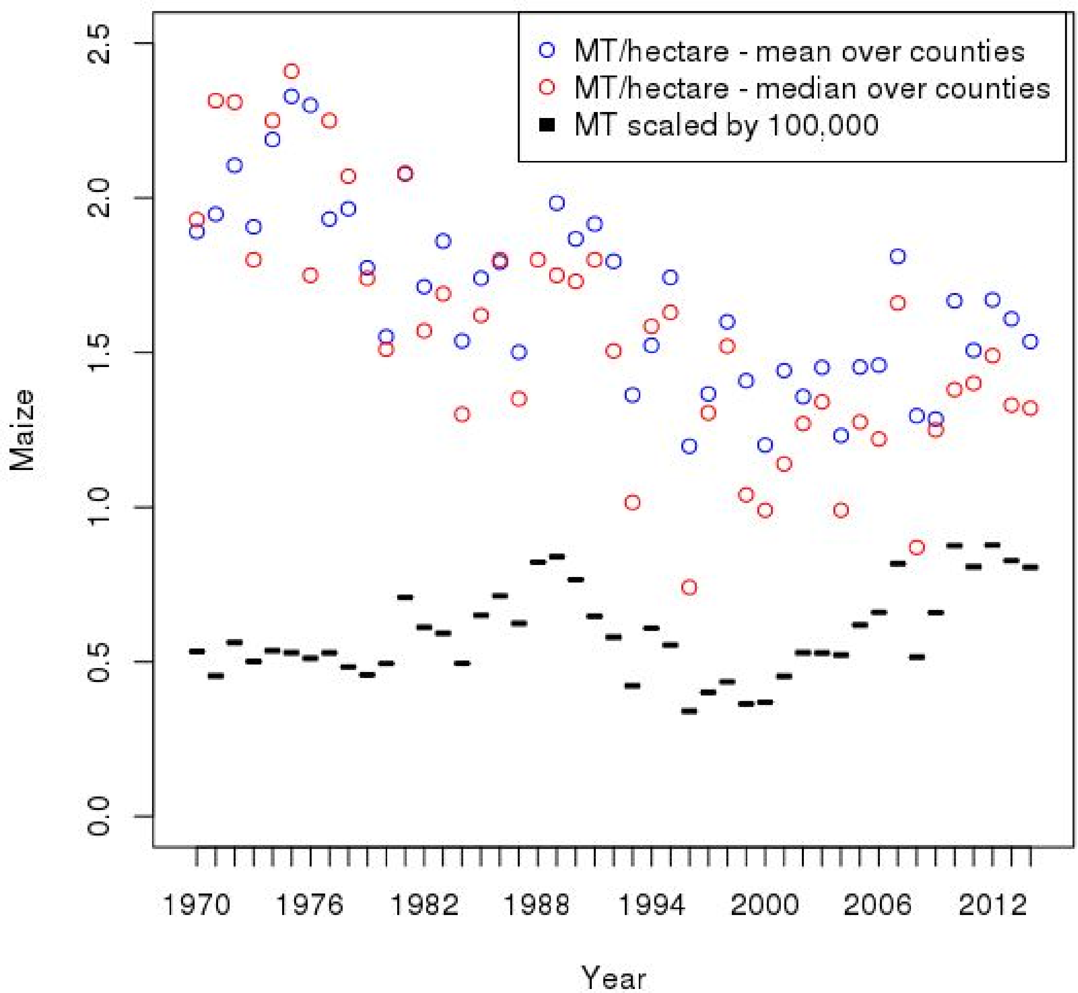
R function *pvcm* from the *plm* package (no AR or MA)

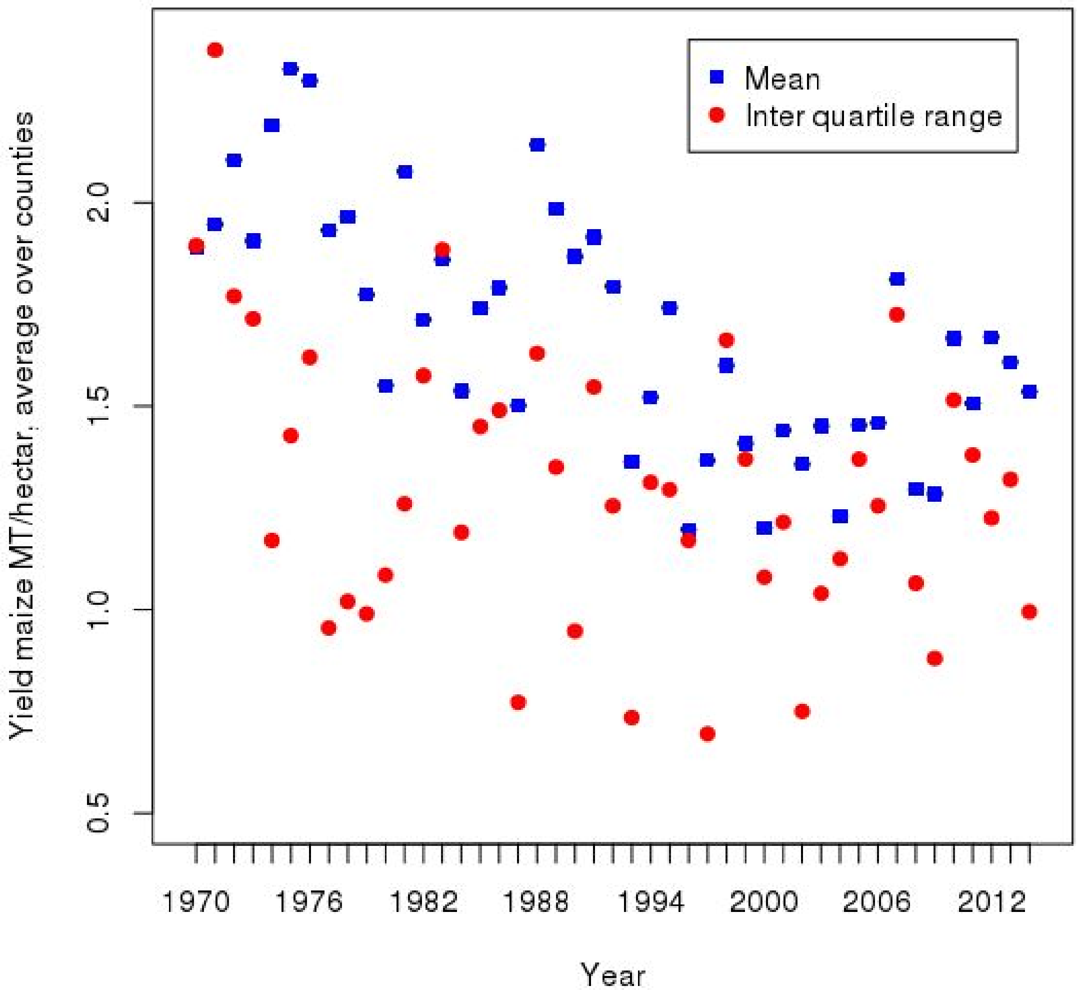
Error structure

Fixed effects	**No AR**	**ARMA(2,2)**	**Swamy (1970)**	**Counties separately (mean)**
Intercept	1.237***	1.036***	1.515***	1.519
Precipitation (Z – score)	0.214***	0.187***	0.218***	0.224
Precipitation – coef. of variation	0.207**	0.143*	0.107	0.198
AIC	1868.164	1817.509		
BIC	1913.900	1881.540		

Descriptive analysis of maize yields data from 1970

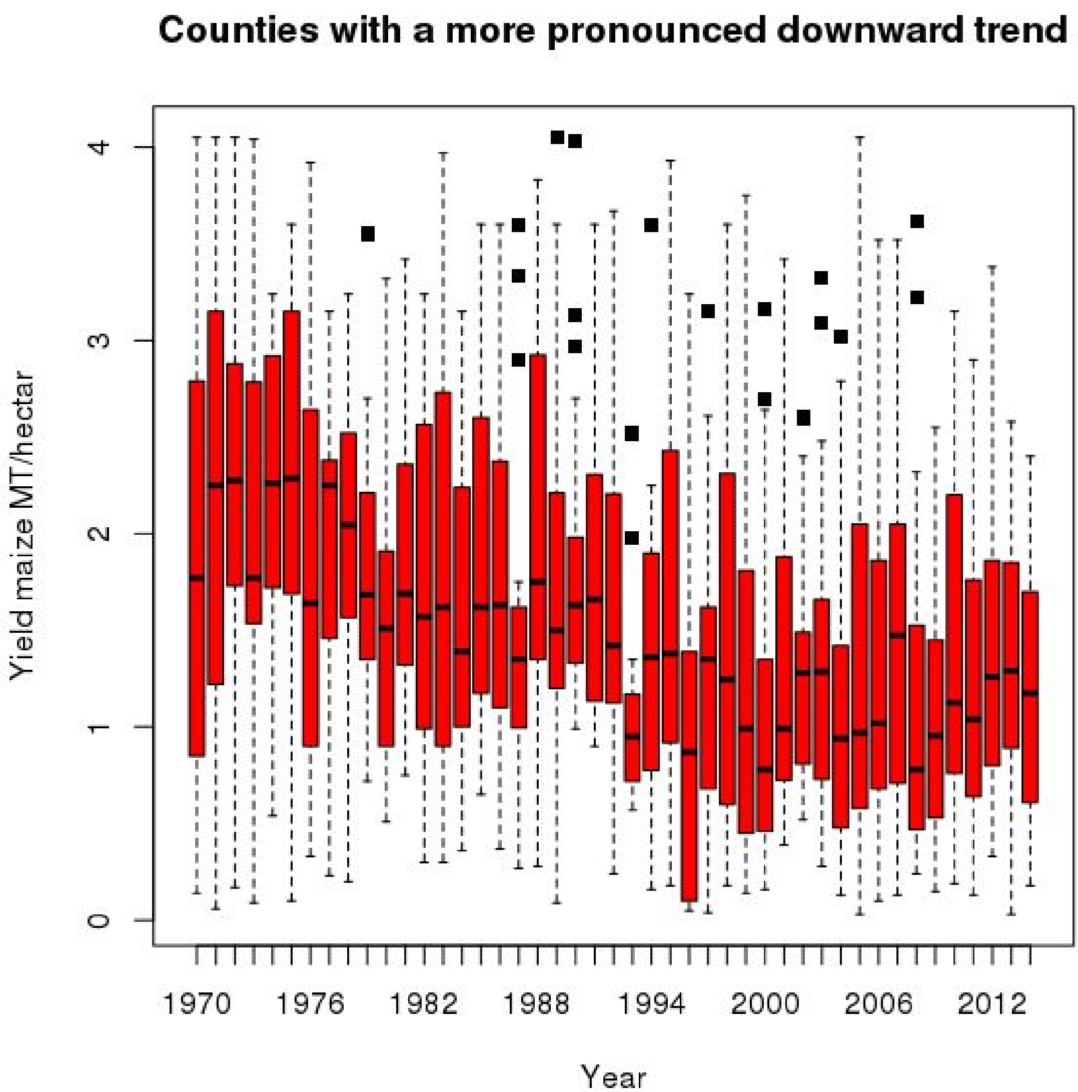
- Interesting point: **a strong downward trend in yields**, at least until 1990. Although the production (MT) exhibit an upward trend.





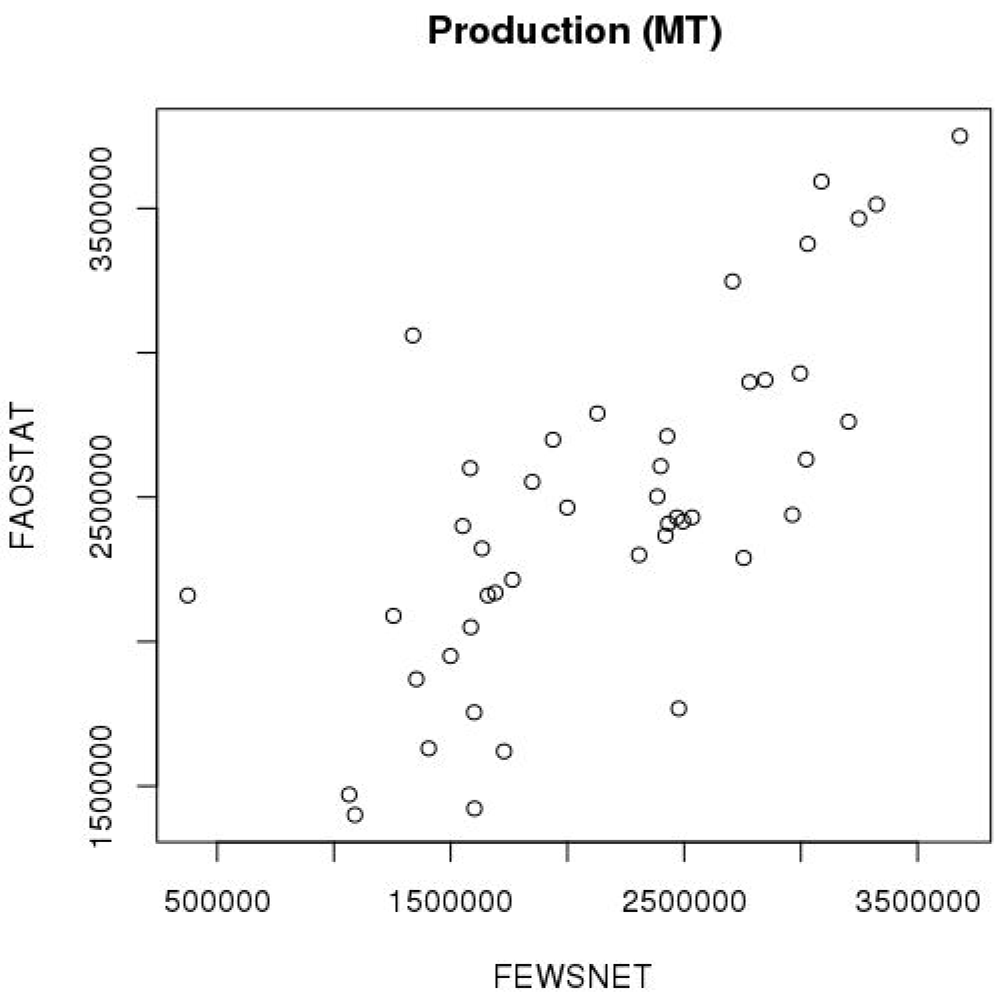
- It appears that the data are more reliable from year 1991 because of the following reasons:
    a. Relatively many n/a's before 1991
    b. Strong downward trend in yields before 1991
    c. The interquartile range appears to be smaller and more stable after 1991
    d. T-tests show that there is a significant difference between the mean before 1991 and after 1991
    e. Before 1991, no separation between short and long seasons. Only one data entry for each county and year. Between 1991 and 2001 there are three data entries for most counties. These are long rains, short rains and together.

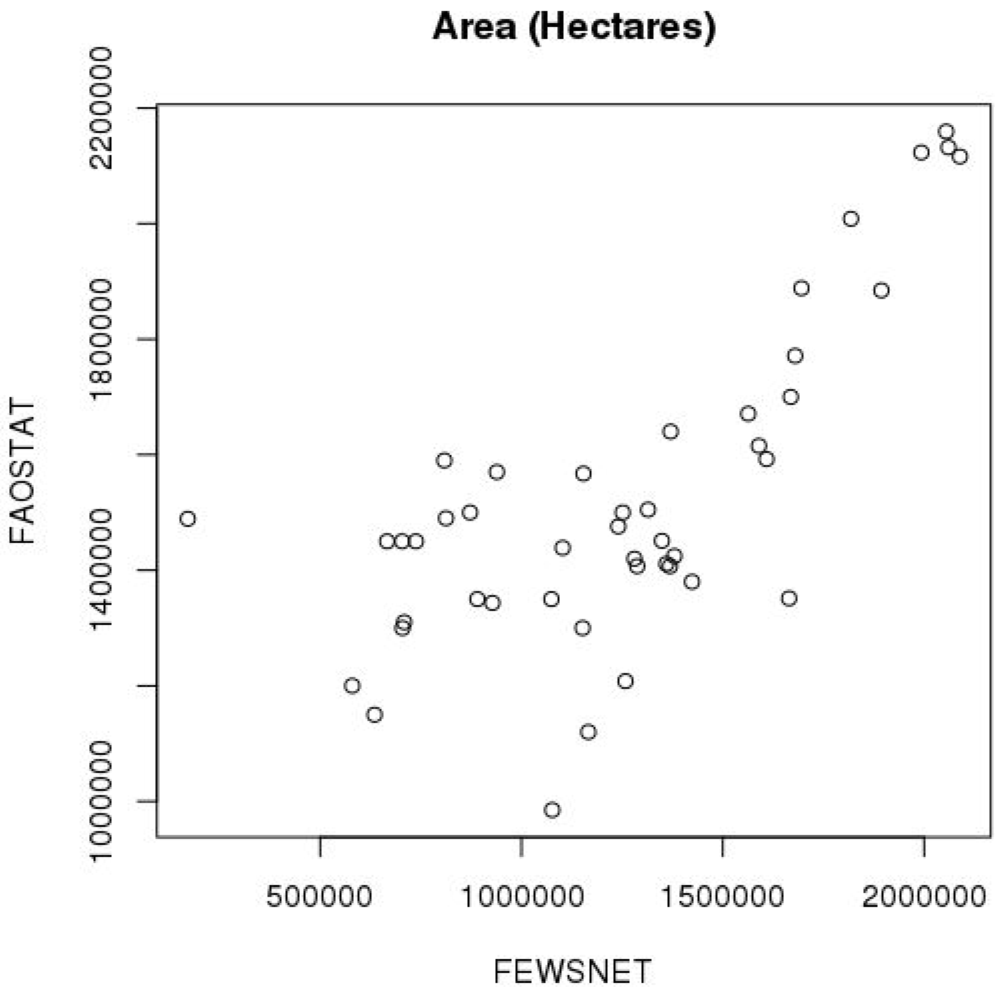
- To investigate, if there are some counties without the downward trend, I plotted each county separately:
    a. <u>Counties **without** (or with much less pronounced) **downward trend**</u>: Nyandarua, Kwale, Lamu, Tana River, Kitui, Mandera, Homa Bay, Migori, Elgeyo Marakwet, Narok, Samburu, Turkana, Uasin Gishu, West Poko, Bungoma, Kakamega , Vihiga , Meru, Tharaka Nithi, Nandi



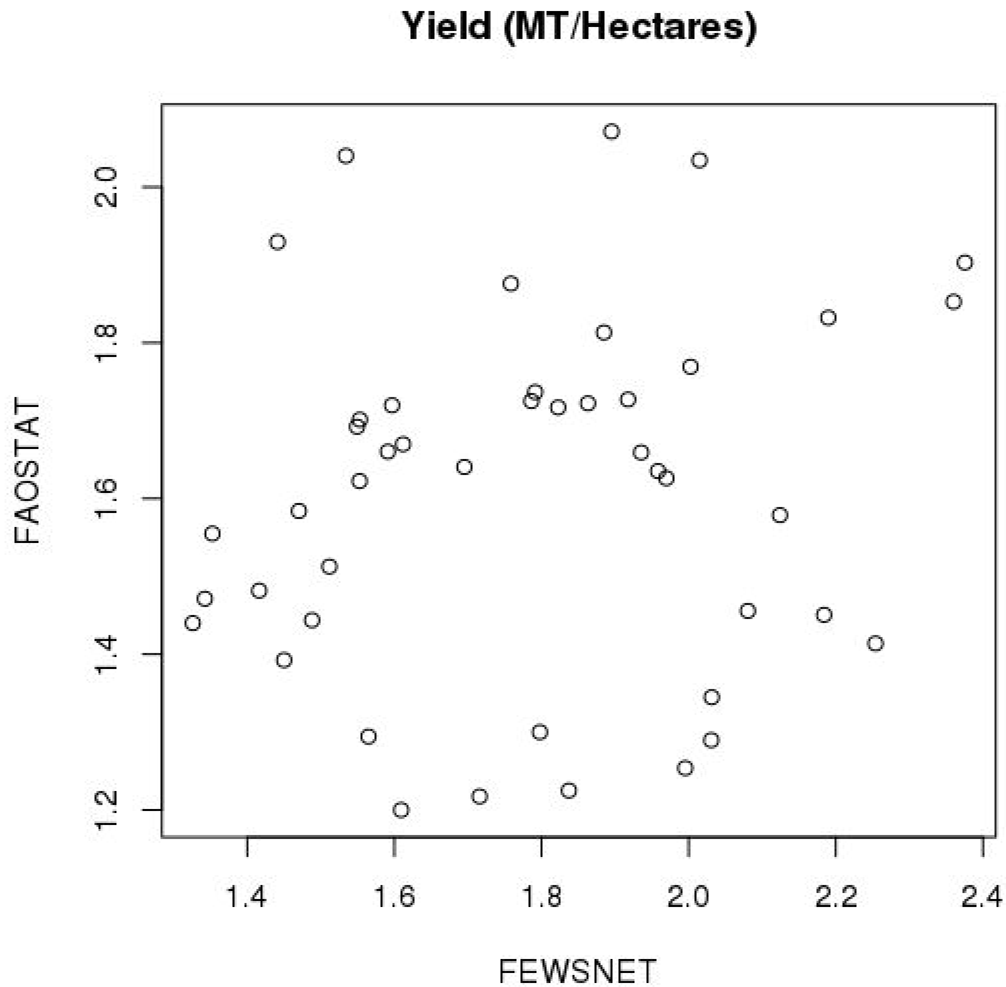
Comparison of the county level yield data (FEWSNET) with the country level FAOSTAT data

- I aggregated the FEWSNET data as follows:
    - For every year, I calculated the sum of production in MT over the counties
    - For every year, I calculated the sum of harvested area in hectares over the counties
    - I obtained the country level yields for every year by dividing the production (obtained as described above) by area (obtained as described above)
    - Then I compared the country level aggregates with the FAOSTAT data





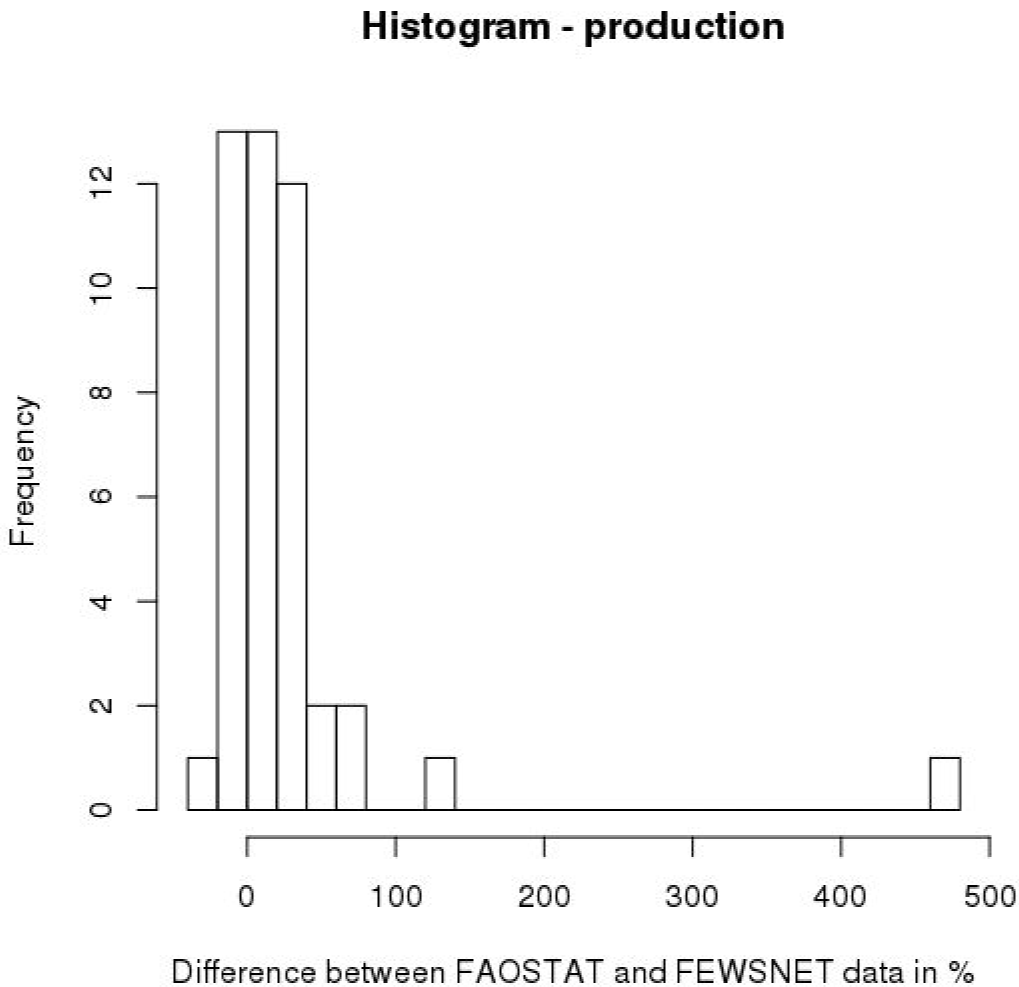
Area (Hectares)
FAOSTAT
FEWSNET
500000
1000000
1500000
2000000
1000000
1400000
1800000
2200000

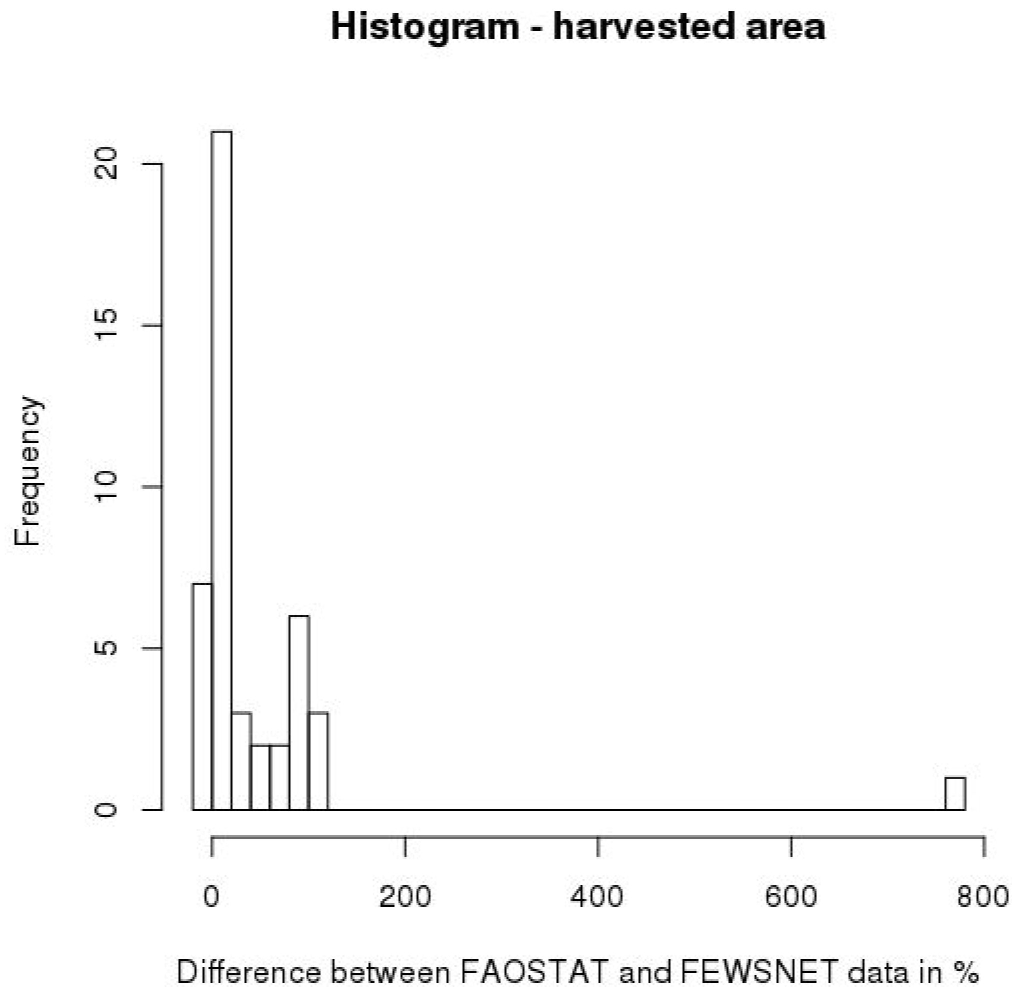


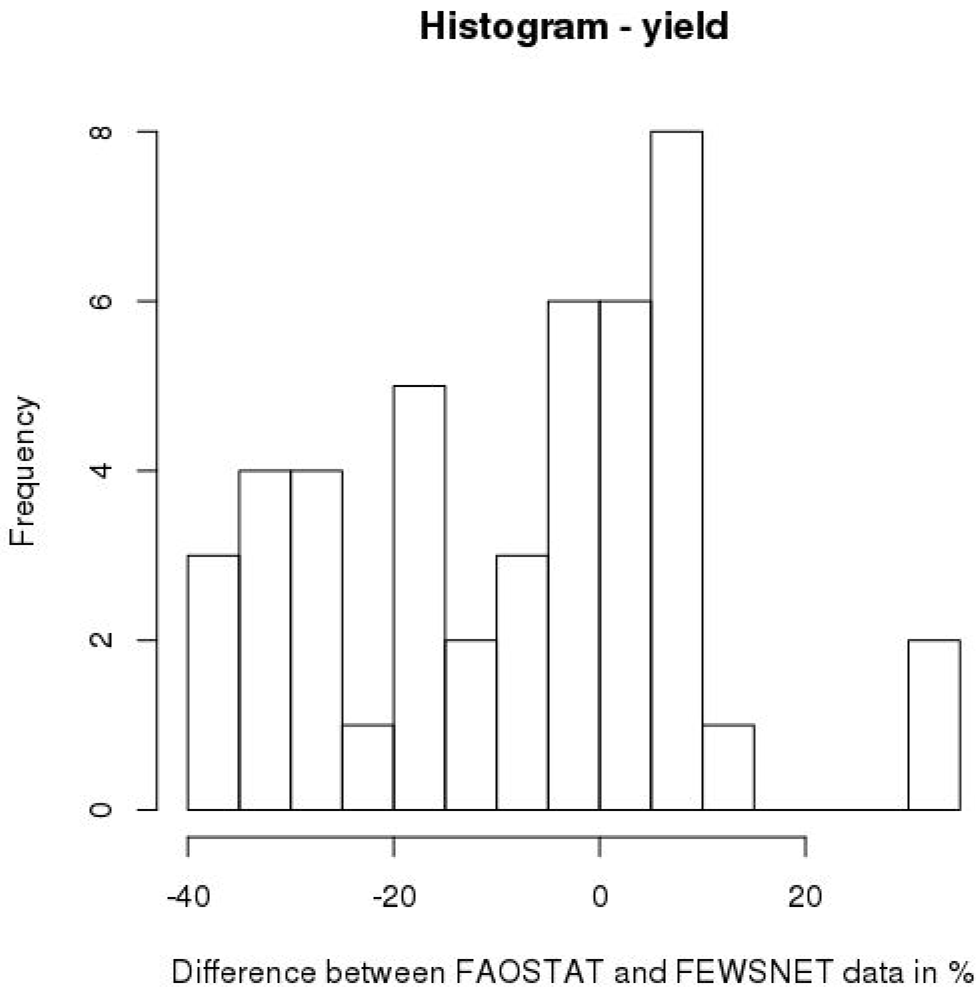
- For each variable (production, area and yield) I calculated relative differences between the datasets in percentage as follows:

  100*(FAOSTAT-FEWSNET)/FEWSNET

- Histograms of the relative differences are below


Histogram - production
Frequency
0
2
4
6
8
10
12
0
100
200
300
400
500
Difference between FAOSTAT and FEWSNET data in %


Histogram - harvested area
Frequency
0
5
10
15
20
0
200
400
600
800
Difference between FAOSTAT and FEWSNET data in %



- Hence, the differences are relatively large for area and production
  - Production:
    - range: -28.6% - 477.2%
    - mean of absolute values: 32.3 %
  - Area:
    - range -18.8% - 769.1 %
    - mean of absolute values: 48.3%

- For yields, the differences are relatively smaller:
  - range: -37.3% - 33.8%
  - mean of absolute values: 15.5 %

- Years with the largest differences:
  - Production:

        - 1996 (477.2%), 1994 (128.6%), 1993 (66.5%), 1976 (64.2%), 1991 (54.6%)
    - Area:
        - 1996 (769.1%), 1973 (117.6%), 1970 (107%), 1975 (106.1%), 1974 (96.5%)
    - Yield:
        - 1975 (-37.3%), 1972 (-37.2%), 1973 (-36.5%), 2004 (33.9%), 1974 (-33.8%)

Climate measures in the literature:

- Erin Lentz et al. (2017): ***prices and consumer strategy index as dependent variables rather than yield***
    - Total rainfall during the last year's rainy season (October to April for Malawi)
    - Timing of the beginning of the rains for the prior year
        - Number of days following the first October when the rains began
        - The beginning of days defined as when it rains for at least three of the past five days for a total accumulation of at least 10 millimeters
    - Dry spells: maximum number of days without rain during last year's rainy season (Oct-April)
    - Maximum daily precipitation that month in regions that are susceptible to floods

- Abraha and Savage (2006)
    - Wet and dry day counts (probably per month)
        - Definition of `wet and dry days' not clear from the paper
    - Monthly total rainfall and its variances
    - Daily and monthly mean and SD of wet day precipitation
    - Min and max air temperatures. Probably daily extremes, these then used for monthly mean and SD (not entirely clear from the paper)

- Adejuwon (2005)
    - Total rainfall for
        i. The first month of the period from sowing to harvesting (June)
        ii. The first two months of the period from sowing to harvesting (June and July).
            - this turns out to be the most important in this case

        iii. The three months of the period from sowing to harvesting (June, July and August)
    - Minimum and maximum temperature of each of the growing season months

- Ben Mohamed et al. (2002)
    - Sea surface temperature anomalies at various locations (see paper for the particular sites) - ***significant***
    - Number of rainy days (details not specified)-***significant***
    - Daily amount of rainfall (as such not usable in our case - not consistent with our yearly frequency)
    - Amount of rainfall in July, August and September - ***significant***
    - Maximum air temperature in the hottest month April (Niger)
    - Minimum air temperature in the coldest month January (Niger)
    - The length of rainy season:
        i. The difference between the dates of the beginning and the end of the season.
            - The beginning of the growing season is defined as being when the amount of rainfall in three consecutive days is at least 25mm and no dry spell of more than seven days duration occurs in the following 30 days.
            - The end of the rainy season is that rainy day after which rain recorded during 20 days is less than 5 mm.

- Blignaut et al. (2009)
    - Rainfall as the annual sum of the provincial monthly average
    - Temperature as the annual averages of the daily maximum temperatures in two or three towns or cities per province
        - The data received as monthly averages of daily maxima, these then used to compute annual averages

- Chipanshi et al. (2003)
    - Daily maximum and minimum temperatures
        - I think that these were used as inputs into a GCM (together with daily solar radiation calculated based on daily sunshine hours). Then climate change simulated->then hypothetical temperature as ***mean monthly values and rainfall as departures from normal*** obtained from GCM under climate change. These then used as intermediate inputs to estimate effects of climate change on yields

- Gbetibouo and Hassan (2005): ***Dependent variable = farm net revenue (Ricardian approach)***

    - *'The normal climate variables based on 30 years average of temperatures and precipitation observed over 1970-2000'*
    - Rainfall pattern (mm)
        - Summer and winter separately
        - Summer and winter months not specified
        - Way of aggregation not specified in more detail, probably cumulative
    - Average temperature (°C)
        - Summer and winter separately
        - Summer and winter months not specified
        - Way of aggregation not specified in more detail

- Giannakopoulos et al. (2009)
    - General circulation model predicts climate change of C2. This used as an input for the CROPSYST model
    - CROPSYST model predicts effects of climate change on crop productivity
        - Measures used as inputs for the CROPSYST (not specified in more detail):
            - Tmax
            - Tmin
            - Total rainfall
    - Other interesting climate measures used to describe climate changes. But probably not used as inputs for the crop productivity model here.
        - Number of summer days defined as Tmax>25°C
        - Number of hot days Tmax>30°C
        - Number of heatwave days Tmax> 35°C
        - Number of tropical nights Tmin>20°C
        - Number of frost nights Tmin<0°C
        - Number of dry days (dry day if daily precipitation amounts to less than 0.5mm = a typical threshold value) and spells
        - Annual maximum running total rainfall over 3 days (potential to cause local flooding)

- Laux et al. (2010)
    - Impact of climate change on attainable yields of maize and groundnut
        - Daily Tmax > 30°C
            - = reference to Rosenzweig and Hillel (1993) who found this for USA, but also
            - reference to Tingem et al. (2008) who found a similar relationship for Cameroon

- Leemans and Solomon (1993)
    - Mean temperature of the coldest month

    - Length of growing period
    - Mean maximum temperature
        - Daytime temperature during growth period would be better, but data not available

- Lobell et al. (2008)
    - Monthly temperature and precipitation

- Lobell and Burke (2010)
    - Temperature - growing season average
    - Growing season total precipitation

- Odingo (1990)
    - Group variable according to Day/Night temperatures??

- Rosenzweig and Parry (1994)
    - Mean monthly changes in temperature and precipitation

- Sagoe (2006)
    - Total rainfall (within year?)
    - Total number of rainy days
    - Highest and Lowest rainfall recorded in the year - in region with the highest (lowest) rainfall

- Schulze et al. (1993)
    - Daily rainfall
    - Daly max and min temperatures
    - Reliable monthly means of daily maximum and minimum air temperature with an altitudinal correction factor (= adjustment for elevation)
    - Degree days with base 8°C from seedling emergence to the end of juvenile stage (Σ°C8-day)
    - Degree days with base 8°C from silking (base 8°C) to physiological maturity (Σ°C8-day)

- Thornton et al. (2009)
    - Monthly values for average daily temperature
    - Average diurnal daily temperature variation
    - Average monthly rainfall (for each pixel)

- Tingem et al. (2008)
    - Thermal time - calculated as growing degree days accumulated through the growing season
    - Average temperature above a base and below a cut-off considered for growing GDD

        - The accumulated thermal time mayu be accelerated by heat/water stress

- Tingem et al. (2009)
    - Thermal time - calculated as growing degree days accumulated through the growing season
    - Average temperature above a base and below a cut-off considered for growing GDD
    - Number of days with temperature Tmax > 30°C (not sure if the correlation with yield is meant to be negative or positive)
        - The accumulated thermal time may be accelerated by water stress

- Walker and Schulze (2008)
    - Degree days Oct-March (this study for South Africa) between 10°C and 30°C
        - Showed in a table of descriptive statistics, not sure if used in model
    - Monthly means of daily max temperatures and daily min temperatures in the summer months (i.e. Dec-Mar
        - Showed in a table of descriptive statistics, not sure if used in model

<u>Summary list of the weather/climate measures that I could use:</u>

- Total rainfall (not sure if they mean for the whole year?)
- Growing season total precipitation
- Total rainfall during the last year's rainy season (October to April for Malawi)
- Total rainfall for
    - The first month of the period from sowing to harvesting
    - The first two months of the period from sowing to harvesting
    - The three months of the period from sowing to harvesting
- Rainfall as the annual sum of the monthly averages
- Cumulative rainfall pattern (mm), summer and winter separately
- Highest and lowest rainfall recorded in the year - in region with the highest (lowest) rainfall
- Monthly mean rainfall (for every month?)
- Monthly standard deviation of rainfall (for every month?)
- Mean monthly rainfall as % departures from the normal (for every month?)
- Annual averages of the daily maximum temperature
- Temperature - growing season average
- Average temperature, summer and winter separately
- Monthly means of daily min and max temperatures (for every month?)
- Monthly SD of daily min and max temperature (for every month?)

- Mean monthly changes in temperature and precipitation
- Monthly average daily temperature (for every month?)
- Minimum and maximum temperature of each of the growing season months
- Maximum air temperature in the hottest month
- Minimum air temperature in the coldest month
- Mean temperature of the coldest month
- Monthly means of daily max temperatures and daily min temperatures in the summer months (i.e. Dec-Mar in South Africa)
- Average diurnal daily temperature variation

Day counts/spells:

- Prior year - Number of days following the first October until when the rains began (beginning of rains= when it rains for at least three of the past five days for a total accumulation of at least 10 millimeters)
- Dry spells: maximum number of days without rain during last (prior??) year's rainy season (Oct-April)
- Wet and dry day counts (per year? or each month?).
    - Giannakopoulos et al. (2009) define dry day as a day when daily precipitation amounts to less than 0.5mm
- The length of rainy season in number of days
    - The beginning of the growing season is defined as being when the amount of rainfall in three consecutive days is at least 25mm and no dry spell of more than seven days duration occurs in the following 30 days.
    - The end of the rainy season is that rainy day after which rain recorded during 20 days is less than 5 mm.

Days above/below a threshold, degree days

- Number of summer days defined as Tmax>25°C
- Number of hot days Tmax>30°C
- Number of heatwave days Tmax> 35°C
- Number of tropical nights Tmin>20°C
- Number of frost nights Tmin<0°C

all above: not sure per what period

- Degree days with base 8°C from seedling emergence to the end of juvenile stage (Σ°C8-day)
- Degree days with base 8°C from silking (base 8°C) to physiological maturity (Σ°C8-day)
- Degree days Oct-March (this study for South Africa) between 10°C and 30°C

Floods:

- Maximum daily precipitation that month in regions that are susceptible to floods
- Annual maximum running total rainfall over 3 days (floods)

The new climate data

CHIRPS, GHCN, Means and medians

The new climate dataset includes data from 65 locations while I was expecting it to include 47 locations (for the 47 counties of Kenya). Most of the extra locations are former districts which were merged into counties in 2010.

The extra counties/locations, which are not in the maize yield dataset are:

ID	NAME
3	**Maragua** Former district, 2010 merged with MUranga
7	**Thika** Industrial town in Kiambu county
17	**Mwingi** Town in the Kitui county
20	**Ijara** Former administrative district. In 2010 amalgamated with Garissa county
23	**Bondo** Town in Siaya county
26	**Kuria** Former administrative district. In 2010 became a part of Migori county
29	**Nyando** Former district in Kenya. In 2010 merged into Kisumu County.
30	**Rachuonyo** Former administrative district. Since 2010 part of Homa Bay county
32	**Suba** Former administrative district. Since 2010 part of Homa Bay county

35 **Buret**
Former administrative district. Since 2010 split between Kericho county and Bomet county

37 **Keiyo**

39 **Koibatek**

45 **Trans Mara**

50 **Nandi North**

51 **Nandi South**

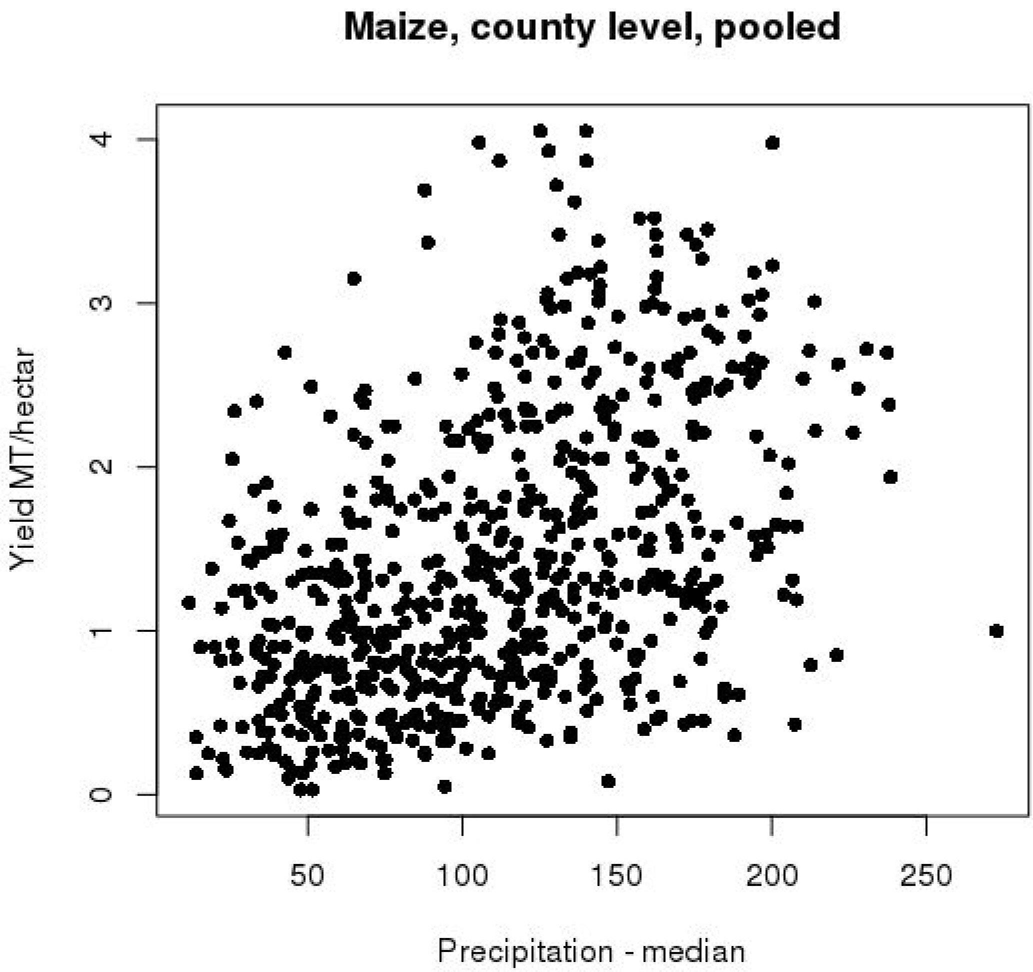
54 **Butere Mumias**

56 **Lugari**

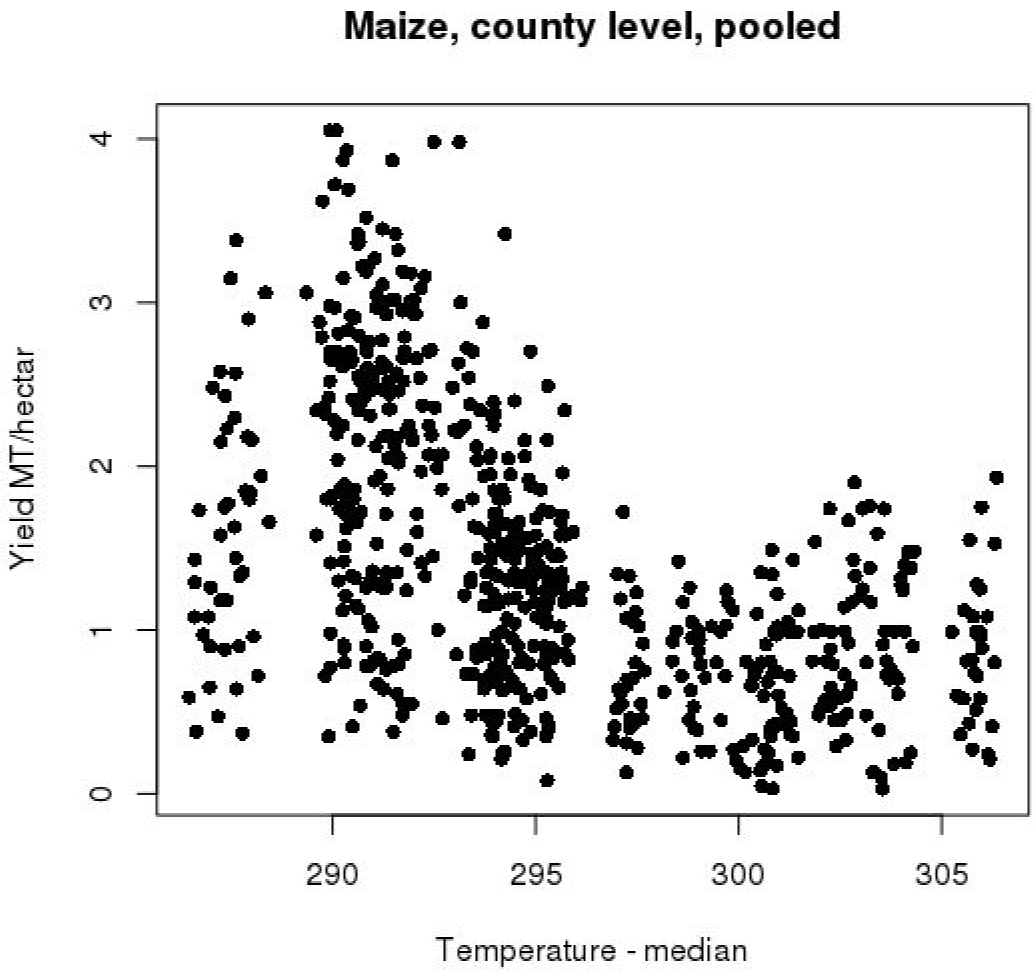
57 **Mt Elgon**

58 **Teso**

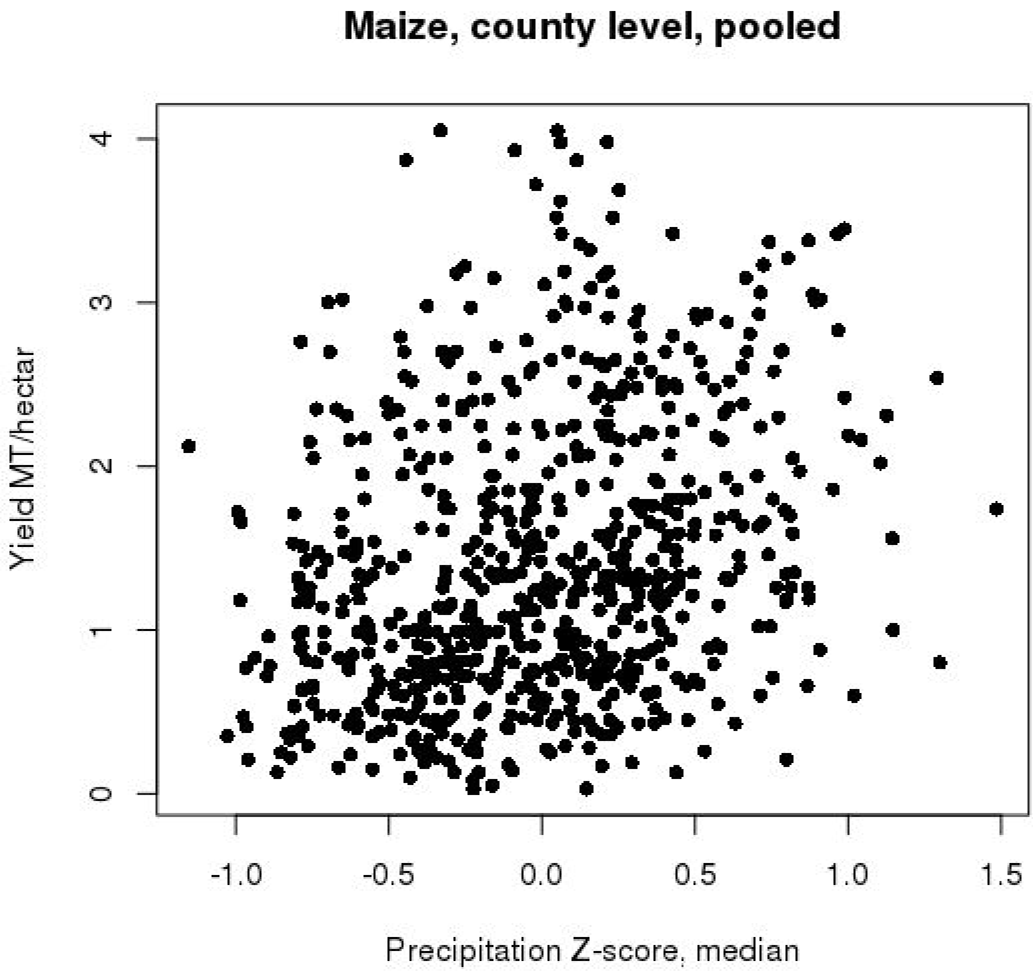
Plots of maize yield against climate/weather data



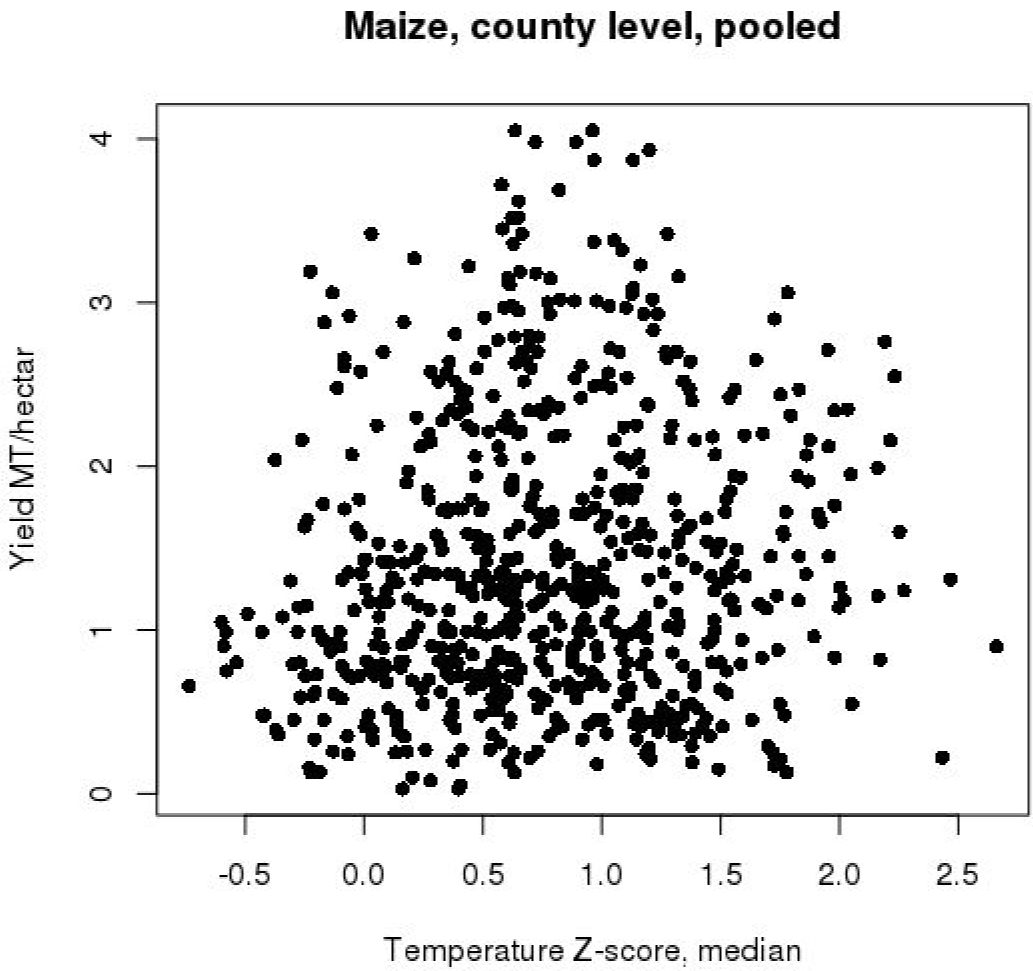
- A nice relationship between Precipitation and maize yield



- An interesting relationship between temperature and maize yields
- Why U-shaped?

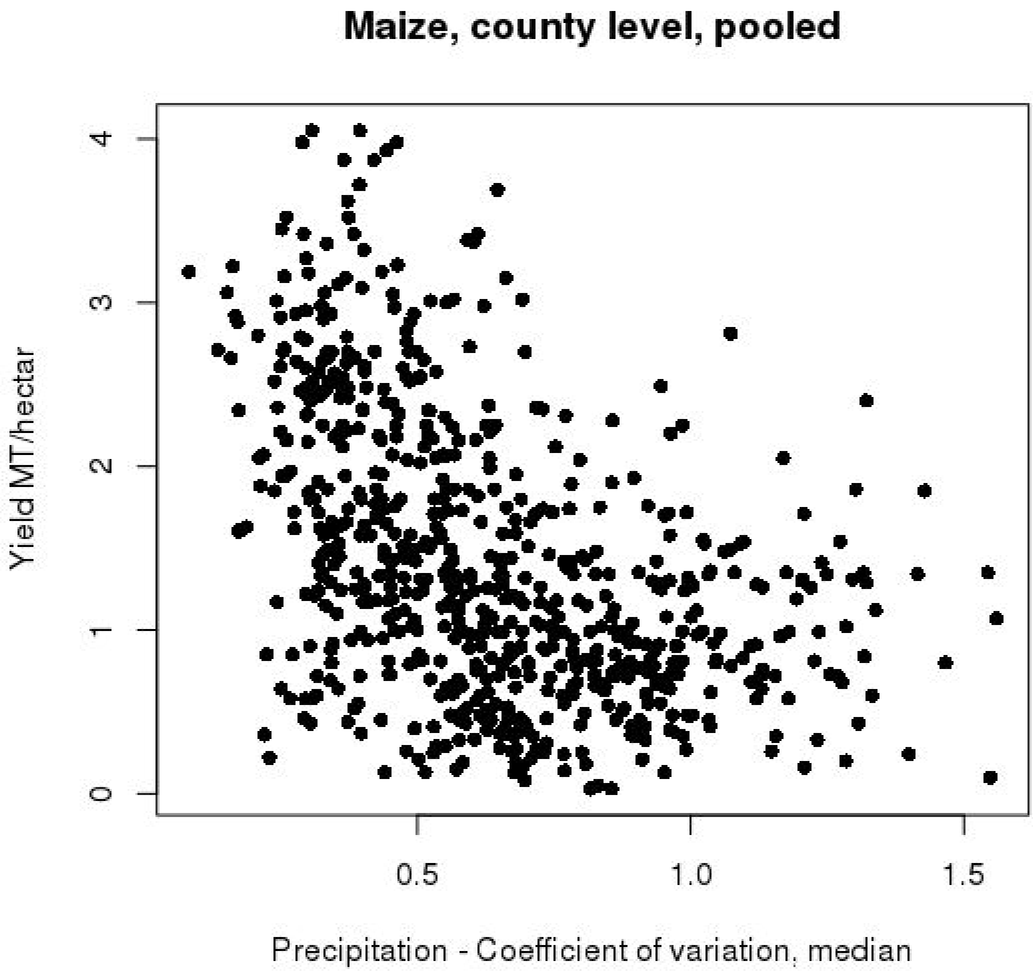


- The relationship between yields and precipitation less pronounced if z-scores of precipitation used instead of the raw data

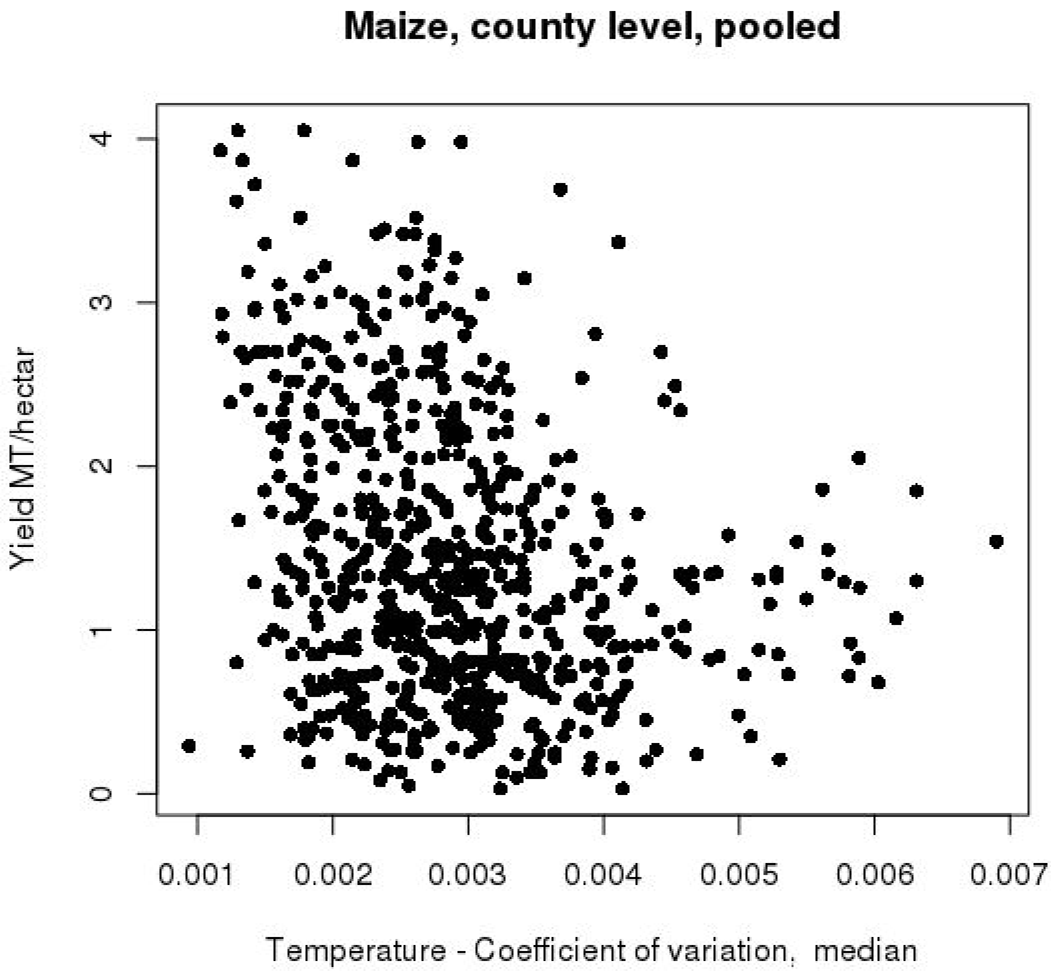


- Also for temperature, the relationship is much less pronounced if z-scores used instead of the raw data

Coefficients of variations:



- There seems to be a strong negative correlation between yields and coefficient of variation of temperature
- This is interesting considering the fact that the coefficients of variation were not significant in my models.
    - Perhaps because I was using a different climate dataset and the Maize outliers were not removed before estimating the models



- There seems to be a negative correlation between yields and coefficient of variation of temperature

Selected specifications of the new mixed-effects models

- The newer version of the climate data
  - Means and medians instead of 90th and 10th percentiles
- Amended definition of the seasons:
  - **Eastern counties:** OND of the last year and MAM of the current year
  - **Western counties:** MAM and JJA of the current year
- Log-linear functional form
- Time period: **2000-2014**
- Yield outliers above 99th percentile removed

$$\ln(Y_{it}) = \alpha + \alpha_i + \beta^{P} P_{it} + \beta^{P2} P_{it}^2 + \beta_i^{P} P_{it} + \beta_i^{P2} P_{it}^2 + \beta^{VCP} vcP_{it} + \beta^{VCP2} vcP_{it}^2 + \beta^{T} T_{it} + \beta^{T2} T_{it}^2 + \beta_i^{T} T_{it} + \beta_i^{T2} T_{it}^2 + \beta^{VCT} vcT_{it} + \beta^{VCT2} vcT_{it}^2 + u_{it}$$

$$u_{it} \sim \mathrm{IID}(0, \sigma_\epsilon^2)$$

P... Precipitation
T… Temperature
Y… Maize yield
t...year
i… county

	Linear functional form			Log linear functional form		
Fixed effects	**Raw data no convergence!!!**	**z-score**	**scaling**	**raw data**	**z-score**	**scaling**
Precipitation			0.220***		0.296***	0.242***
Precipitation squared			-0.045º		0.008	-0.050º
Precipitation - c. of variation			-0.063*		0.0006	-0.003
Precipitation - c. of variation squared			0.001		0.0000006	-0.001
Temperature			-0.603***		0.067	-0.500***
Temperature squared			0.229**		-0.024	0.102
Temperature - c. of variation			0.030		0.001	-0.023
Temperature - c. of variation squared			-0.010		-0.000002	-0.019
REML criterion			1065.8		1099.9	1013.