

# Modeling Impacts of Drought on all Major Crops in Ethiopia

*Michael L Mann & James Warner*

*September 29, 2017*

## Introduction

The ability to monitor and predict crop yields in developing countries is critical to the successful adaptation to changes in our climate. Increased temperatures and variability has already been linked to losses in maize and wheat yields (-3.8 and 5.5% respectively) and crop prices globally (Lobell, Schlenker, and Costa-Roberts 2011). Although much effort has been placed on modeling the spatial distribution of these shifts, less effort has been placed on how yields vary across space and time (Ray et al. 2015). Advances in remote sensing provide avenues to monitor agricultural crop health at high spatial and temporal resolution. However, our ability to monitor changes in plant productivity is still limited in the more complex environments common to many developing countries (Mann and Warner 2015).

Remote sensing based efforts to characterize the extent, cultivation practices, and productivity of global croplands has a long history. In fact, agricultural monitoring motivated much of the earliest work in remote sensing for example NASA's LACIE and AgRISTARS programs in the 1970s and 1980s and (MacDonald and Hall 1980, Hatfield (1983), NASA (1984), Pinter Jr et al. (2003))). Since then, substantial progress has been made in mapping cropland extent, crop types, irrigation status, cropping intensity, and productivity from remotely sensed imagery. For example, the MODIS Land Cover Product MCD12Q1 (Friedl et al. 2002, Friedl et al. (2010)) provides operationally produced, global scale maps of agriculture and agricultural-natural mosaics at an annual time step and 500 m spatial resolution from 2001-present. A finer resolution (~30 m) data set is available for the conterminous United States which maps the annual extent and type for over 250 crops using primarily Landsat imagery: the Cropland Data Layer [NASS (2003)]. These are but two prominent examples out of a broad literature documenting a wide variety of efforts to map cropland extent and type from remotely sensed imagery (Lobell and Asner 2004, Xiao et al. (2006), Thenkabail et al. (2007), Ramankutty et al. (2008), Wardlow and Egbert (2008), Biradar and Xiao (2011)). Remotely sensed imagery has also been employed to map irrigated areas (Thenkabail et al. 2009, Portmann, Siebert, and Döll (2010)), and cropping frequency/intensity (Biradar and Xiao 2011, Gray et al. (2014), Li et al. (2014)).

Initial efforts (e.g. LACIE and AgRISTARS) primarily utilized remotely sensed imagery to characterize the spatial extent and growth stage of crops, but relied on models driven chiefly by meteorological information to predict crop yield (Idso, Jackson, and Reginato 1977, Doraiswamy et al. (2003)). However, the biophysical link between canopy spectral reflectance and net primary production has long been established (Tucker and Sellers 1986); indicating that satellite measurements could play a role in determining crop yield directly. Indeed, early experimental work confirmed the usefulness of spectral measurements in predicting LAI and intercepted PAR in crops (Daughtry, Gallo, and Bauer 1983, Asrar et al. (1984), Clevers (1997)), a result that was later extended to satellite measurements of spectral reflectance (Tucker et al. 1980, Grotewold (1993), Bartholome (1988)). Spectral measurements typically explain variability in LAI and intercepted PAR better than crop yields because a variety of factors other than net primary production (e.g. weather during critical crop growth stages) influence yield. Nevertheless, a wide number of studies have documented highly explanatory empirical relationships between satellite measures such as NDVI (in many forms: growing season maximum and mean, seasonally integrated, etc.) and yields for a variety of crops, particularly at regional scales (Rasmussen 1992, Benedetti and Rossini (1993), Funk and Budde (2009), Becker-Reshef, Vermote, et al. (2010), Becker-Reshef, Justice, et al. (2010), Mkhabela et al. (2011)). Because certain crop growth stages are particularly critical for final yield (Butler and Huybers 2015), improved results are often seen when remotely sensed data are used to characterize crop phenology (Bolton and Friedl 2013). More recently, methods for forecasting yields with remotely sensed variables at the field scale have been explored (Lobell et al. 2015, Gao et al. (2017)).

In addition to establishing a direct relationship between satellite measurements and crop yield, combining these observations with model output, through formal or ad hoc data assimilation techniques has also been demonstrated (Mo et al. 2005, Moriondo, Maselli, and Bindi (2007), De Wit and Van Diepen (2007)).

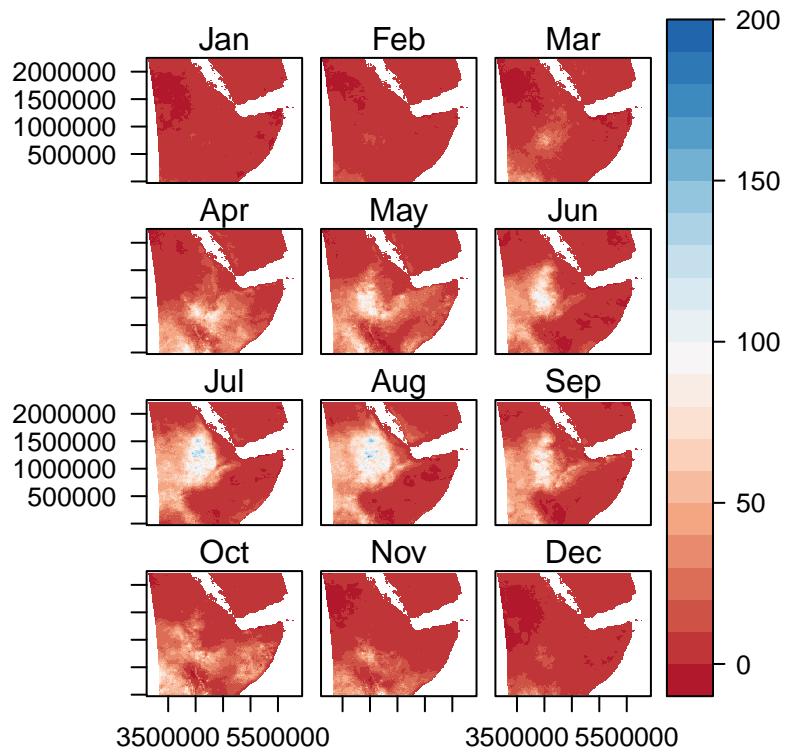
The main objectives of this paper is to create a suite of models that can a) provides reliable estimates of yields based on remotely sensed data, b) reveal insights into the impacts of management. We also create a suite of algorithms used to extract, summarize, and organize remotely sensed data and prepare it for spatiotemporal analysis.

## Drought meets distribution

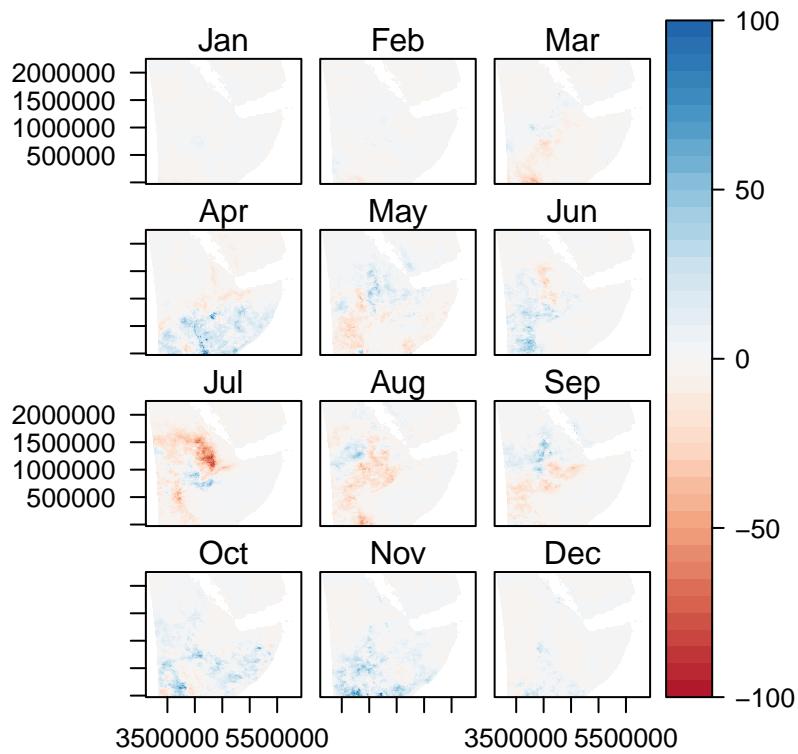
Spurred by a long history of monsoon failures across the Sahel, forecasters and aid teams anxiously watched eastern Africa at the onset of a significant El Nino event in late 2014. Reports of below average rainfalls in 2015-17 across the low lying pastoral communities and South Eastern Ethiopia triggered an international effort to provide food aid to roughly 8.5 million people in the summer of 2017 (“Ethiopia: Drought - 2015-2017,” n.d., (“Mid-Year Review, Ethiopia Humanitarian Requirements Document, July 2017” 2017)). The understated response from the Ethiopian government has lead to accusations that the government is downplaying the severity of the drought as early as 2016 (Schemm 2016, Schemm (2017)). Despite the reality of a severe and crippling drought in food insecure low lying areas, there is little sign of food shortages nationally, with a recent study observing no significant change in grain prices throughout the country (Bachewe, Yimer, and Minten, n.d.). The lack of a price response would indicate that while food insecure areas were hit heavily by water shortages largely effecting livestock, while leaving the highland areas largely unaffected or triggering only minor losses. Despite international concern, conditions up until the 2016-2017 growing season may have been largely handled by domestic resources and trade flows, therefore requiring shifts in distribution rather than a large-scale international intervention.

Looking at Figure 1 we can see the spatiotemporal patterns of rainfall across Ethiopia. The top panel (a) reports the median total rain fall per 10 day period during each month (in mm). Here we can see the onset of the heavy rains during the Meher growing season, and the decline near harvest in October. In panels (b-c) we can see the deviation from this norm for the years 2015 and 2016 (in mm). The only striking shortfall is July in 2015 and patchy isolated declines through August that year. 2016 by comparison shows indications of generally above average rainfall. These figures also demonstrate the spatial heterogeneity of rainfall over time. Both years show small areas that interchange between having above or below average rainfall. It should be pointed out that the erratic nature of these rains might point to the importance of the timing of these rains, with rainfall being more important during critical growth stages like germination and flowering. If rains fail during critical periods, even small shortfalls can precipitate significant losses. Therefore Ethiopia, like many rainfall dependent countries has and will continue to be be prone to localized and occasionally wide-spread yield losses.

**(a) Median 2009–2015**



**(b) 2015 Deviation**



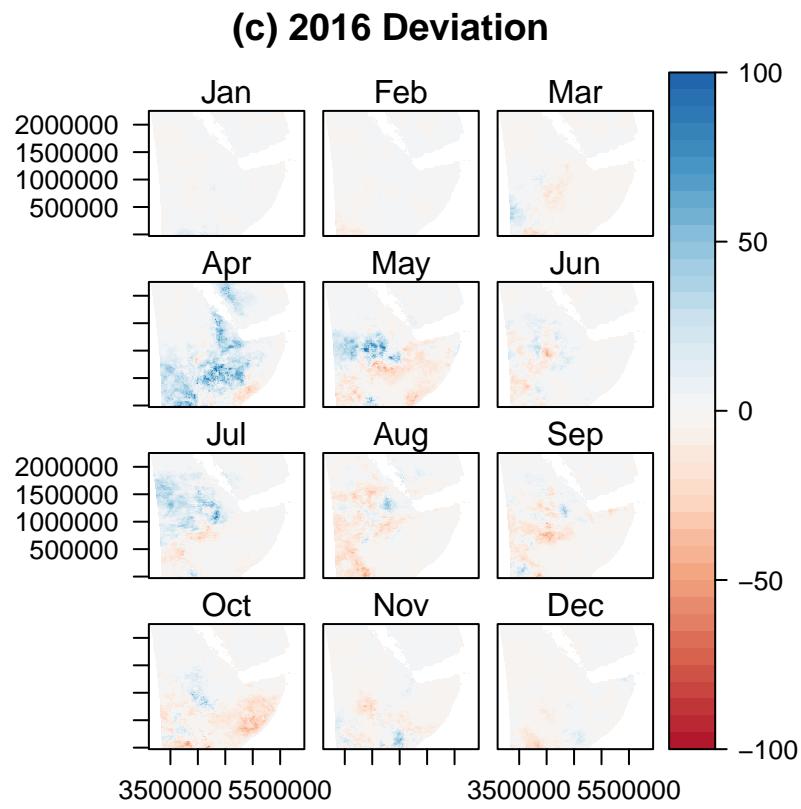


Figure 1: Median rainfall per 10 day period 2009-2015 compared to 2015 and 2016

Looking at the example of wheat production for the top 15 producing zones (Figure: 2), we can see that total output has remained steady or increased in the majority of top performing zones. This can be seen by the increase in total production from 49,500 in 2010 to 64,209 in the 2015 growing season. Additionally, we can see that the total production of each zone (as indicated by the width of each bar) generally increases over time, or remains relatively constant. Moreover we can see that the rank of regions (as indicated by the plot order, bottom being lowest rank, top being highest rank) remains surprisingly steady despite the exogenous shock of drought beginning in the 2015 planting season.

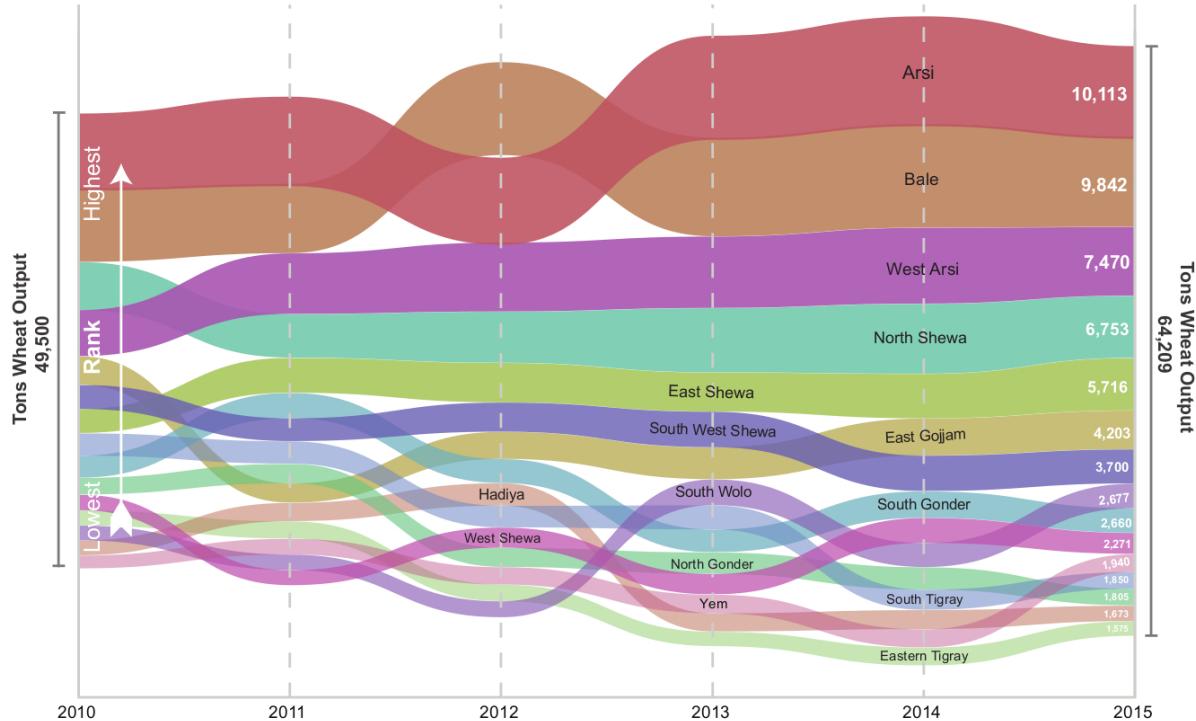


Figure 2: Total wheat output of top 15 producing zones in tons

We can see from Figure 3 three time series for the major regions depicting the drought, in the top panel (a) the percentage of planted area damaged by damage type, the middle panel (b) the percentage of area planted in each major crop, in the bottom panel (c) median output per hectare (OPH) in quintiles for each major crop. Although we see a marked spike in reports of drought damage in the 2015 planting season (covering harvest in 2016), we see a few important responses. First, we see that although slight, farmers increased the percentage of area planted in drought tolerant crops such as maize and sorghum. Second, we see that although some regions saw declines in median output per hectare, these losses are largely offset by gains in other crops or by the steady increase in OPH since 2010.

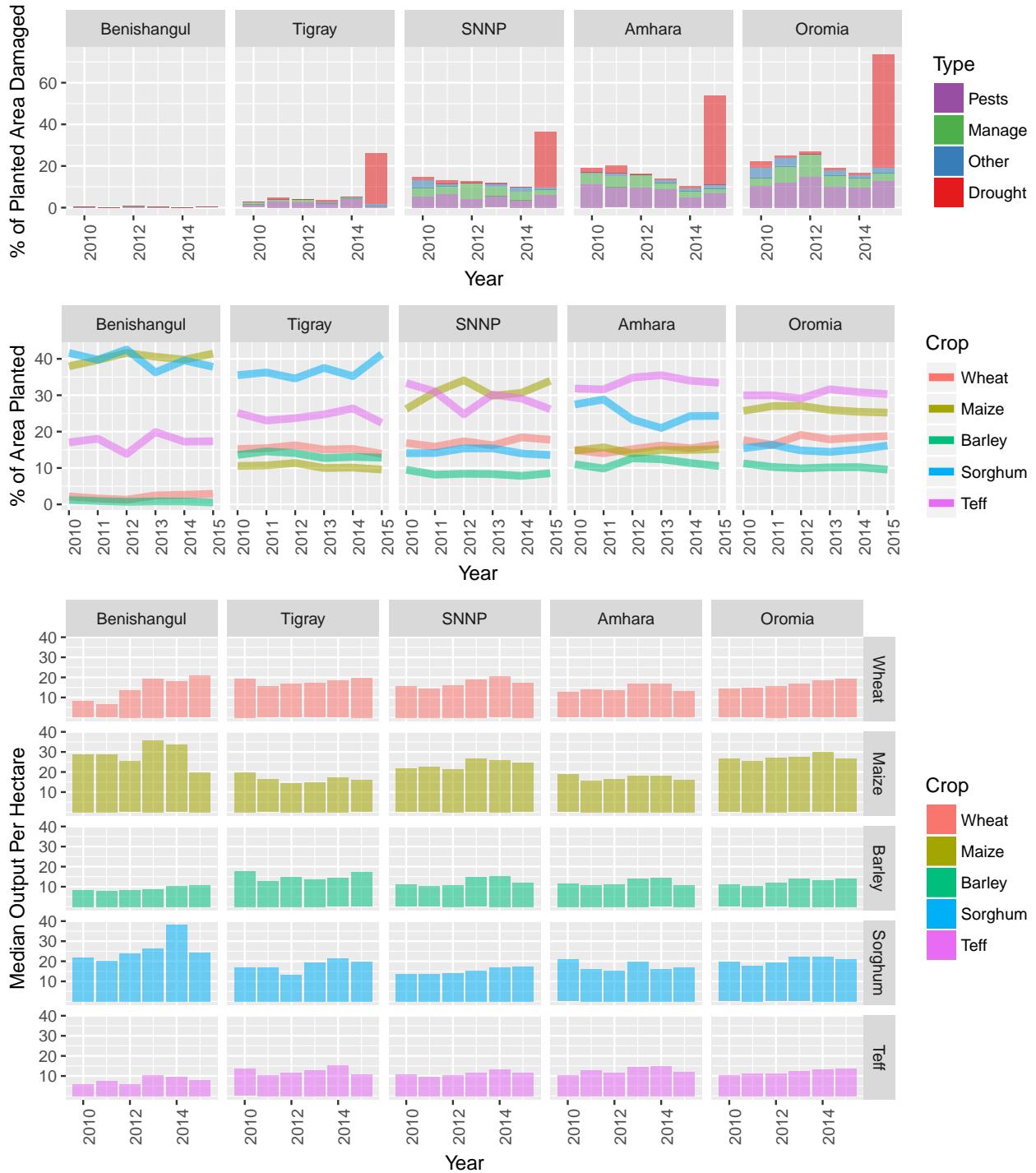


Figure 3: Drought in one graph

Moreover Ethiopia has rapidly improved its ability to mitigate food insecurity through trade flows. International and international trade has likely been improved by the rapid development of the paved road network, and with the opening of the Ethiopia-Djibouti electric railway line in 2016 improving trade with this red sea port. Even in 2011, for the case of wheat, we can see how imports and internal trade flows move wheat from surplus areas to deficit areas. In Figure 4, we see large international imports as well as comprehensive internal trade flows at the zonal level.

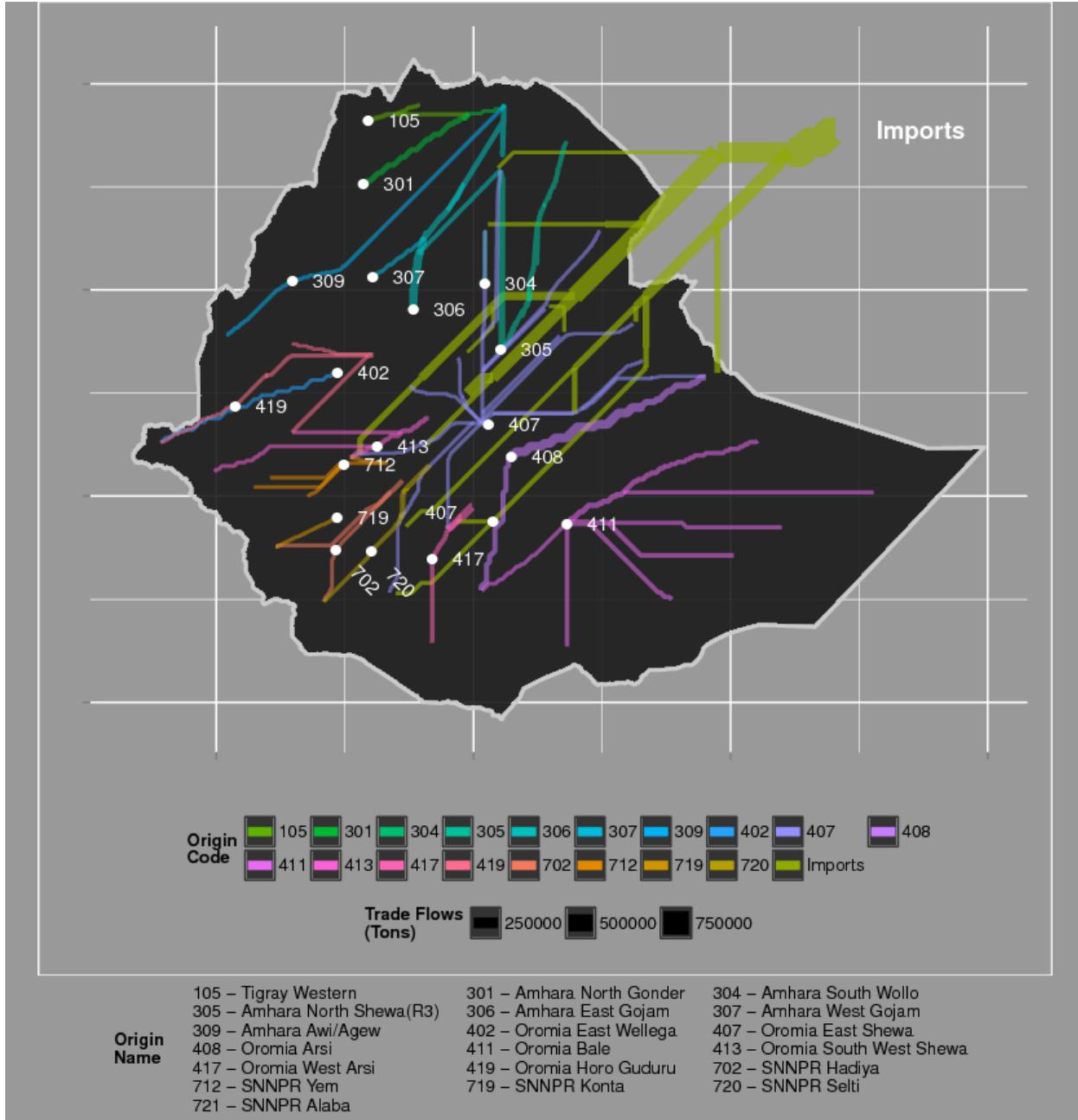


Figure 4: 2011 aggregate wheat trade flows at the zonal level

Although far from perfect we should acknowledge substantial improvements in Ethiopia's ability to mitigate and respond to localized drought.

# Methods

## Data Sources

### **Survey data - Agricultural Sample Survey Dat (2010-2015)**

Survey data was obtained from Ethiopia's Central Statistical Agency's (CSA) Agricultural Sample Survey (AgSS) and was chosen for its annual collection, spatial coverage, and unique sampling structure through these six crop years. The AgSS is an annually collected government administered large-scale survey tasked with measuring agricultural production in Ethiopia at the zonal level.<sup>1</sup> The survey interviews approximately 45,000 farmers per year on a range of farm management questions covering some basic demographics of the household as well as a range of questions concerning planting, harvesting and selling at the plot level. Typically, about 20 farm households are randomly sampled from small local village-level areas of approximately 200 households (the sub-kebele level). On average, based on population, approximately five of these sample areas exist at the kebele-level administration area. Although the figures vary widely, sub-kebeles are meant to approximate about 200 households per unit and the average kebele has around 1,000 households each. From this sampling frame, a random selection of about 2,200 sub-kebeles are chosen as a representative sub-sample for zonal level production projections of major agricultural production areas. Population weights are then applied to project agricultural production at the zonal level. For this study, we construct a panel data set over the six crop seasons identified because approximately 75% of the same sub-kebeles were sampled by CSA over the six meher crop seasons of interest. This effectively builds a base of five relatively favorable crop seasons, from 2010 to 2014, and allows for the drought effects of 2015-2016 as a study in contrast in terms of this significant weather event. The sample is with replacement and therefore representative of the sub-kebele but not a true panel because the same farmers were not chosen. In addition, no sampling weights are applied because we have omitted approximately 25% of the sample and are generally most interested in the localized effects.

While this study collected data on the five principal Ethiopian field crops (teff, wheat, maize, barley, sorghum), because of their predominant importance in value terms for Ethiopia as well as their geographically wide-spread adoption, the analysis should therefore be placed within this context. This study covers the drought effects of only these five crops and not all agricultural production. In other words, when we outline percentages of crop losses we are referring to losses of only these five crops and not total output in the country. We have no reason to believe that these are not generally representative of all crop production, in terms of the drought, but a more detailed analysis would be needed to make this assertion.

The principal unit of analysis is the sub-kebele and all relevant CSA data is aggregated to this level. The survey data represents an amalgamation of all 20 households as a single representative farmer, we refer to as a "super-farmer." This was done for a variety of reasons including the fragmented plot farming system common in Ethiopia as well as CSA data collection methodology. More specifically, CSA data collection methodology relies on crop cuts to estimate productivity at the local level. Depending on the actual number of farmers growing the specific crop, CSA collects up to five different individual farmer crop cuts, averages the yields, and projects this figure onto all plot areas for that crop in the sub-kebele. In this way, individual plot level yields are not able to be determined and, consequently, household level estimates are not possible. Additionally, while the level of the pixel used for the remote sensing data is relatively granular ( $6,250 \text{ m}^2$ ), cost and logistic considerations make plot level estimations prohibitive in this research. Therefore, the data from 20 farmers is aggregated to a single super-farmer resulting in 1,780 observations sampled from the sub-kebele area (approximately  $23,900,000 \text{ m}^2$ ) over a six-year period. The total number of observations collected is about 10,680 but will vary in sample size depending on the crop choice (i.e.. some areas do not grow all five designated crops). A breakdown of each super-farmer is done with aggregated non-crop related information combined with the specificity of the individual crop chosen. More precisely, several variables are consistent across all five crop choices (elevation, proportional numbers of male and female household heads,

---

<sup>1</sup>Beyond the nation, Ethiopia has four official levels of administrative areas. These include, in order of geographic size, regions, zones, woredas and kebeles. There are approximately 12 regions, 88 zones, 690 woredas and 15,000 kebeles. For our purposes, we use the sub-kebele, not recognized as a geographic region for administrative purposes, but commonly used for statistical sampling and are roughly based on population. There are over 75,000 sub-kebeles in the country.

total land area, etc.) and others are crop specific (use of fertilizer, seed, etc.). Each crop year was checked for consistency, cleaned and aggregated to the sub-kebele level. The six years were then aggregated and merged with other data to develop the panel data set. Importantly, for our analysis, we chose specific variables to reflect farmer management decisions and other actions affecting yields. For data cleaning, we Winsorize yield estimates capping them at the 95th percentile, and we also drop any sub-kebele unit that has less than 4 years of data from regressions.

### **Remotely sensed data - Greenness, precipitation and evaporation**

#### **Normalized Difference Vegetation Index**

Considering the relatively small scale of agriculture in this region we use 250m vegetation products from the MODIS satellites. Vegetation indices are obtained from two 16-day MODIS products (MOD13Q1, MYD13Q1) from the Aqua and Terra satellites (Didan and Huete 2006). Due to the staggered nature of acquisition, these products are treated as partially overlapping windows representing 8 day periods (Doraiswamy, Stern, and Akhmedov 2007). We find that the combination of these two products provides a stable and informative time series.

For the sake of simplicity and replicability, we focus on the Normalized Difference Vegetation Index (NDVI) and examine its predictive power using panel econometric techniques. NDVI is sensitive to the amount of chlorophyll in any given pixel and is commonly used to estimate plant productivity and health in agricultural applications (Mann and Warner 2017,Mann and Warner (2015),Funk and Budde (2009)). After removal of snow, cloud and other flagged low quality cells, we remove all non-agricultural cells through the use of the 500m MODIS land cover product (MCD12Q1) for the appropriate year.

#### **Water Availability Variables**

In order to estimate the effects of precipitation (ppt) on our crop production models the Climate Hazards Group Infrared Precipitation Station (CHIRPS) data was included. Data is collected as total precipitation by dekad (Funk et al. 2015). A dekad represents three periods each month, the first two periods being 10 days each, and the third being any remaining days of that month.

Hydrological availability of water and available energy for plant growth have been shown in previous research to be important factors in crop models and are therefore included here. For this reason, we include monthly estimates of potential evaporation (PET) and actual evapotranspiration anomaly (ETA). Actual evapotranspiration (AET) is the sum of transpiration of water through plants plus evaporation from soils and water surfaces. AET is a correlated with vascular plant productivity and correlates with the biomass accumulation and regrowth(Major 1967,Mann et al. (2016)). The ETA variable used in this study is the current AET compared to the 2003-2013 mean. ETA therefore is a proxy for drought as low current values of AET would correspond to reductions either in available energy to move water, or water itself. Potential evapotranspiration is an estimate of the total amount of water that could be moved through the system if water was not a limiting factor. As such it is an ideal indicator of available (or excessive) energy in the form of light and heat, but also includes the impact of wind speed, pressure, and humidity amongst others. The standard deviation of PET, reflects variability at the local level of these weather conditions. Both ETA and PET are available for download through FEWS NET (n.d.).

### **Estimating Crop Yields**

#### **Preprocessing**

#### **Data Aggregation**

Because historical crop yields are only available at the level of certain political units (EAs in this case), pixel-level NDVI data must be aggregated at this same level. In this study, we calculate the mean across the

raw NDVI values for all available agricultural pixels at the EA-level and further extract statistics of interest from this resulting time-series (at a frequency of 8-days).

### **Summarizing Temporal Data**

One of the primary challenges in utilizing high-frequency time series (8-day NDVI) to estimate low-frequency events (seasonal yields) is in reconciling this temporal mismatch when formalizing the relationship between sources of data. In general, low-frequency properties must be extracted from the high-frequency time series that may be relevant to characterize and identify important aspects of plant phenology across the growing season that affect crop yields. In this paper we aim to capture relevant phenological features of wheat through 41 metrics summarizing 8-day NDVI data. These metrics cover Meher growing season statistics, spanning the estimated planting date until harvest date. These plant and harvest dates are estimated on a EA by EA basis, therefore differences in timing due to elevation, crop choice or other management considerations should be captured.

We estimate three classes of statistics across each of these two distinct periods: summary statistics (e.g. mean, max, variance); integrated summary statistics (e.g. area under the curve for the first half of the growing season); and deviation from the norm statistics (e.g. deviations of a given statistic from its 95th historical percentile).

### **Estimation Strategy**

#### **Variable Selection**

One of the key challenges is how to decide on which of the 126 variables from AgSS and MODIS we want to include in the model. For this task we relied on a new suite of variable selection tools coming out of the computer science domain. We utilized VSURF that utilizes a two-stage strategy based on preliminary ranking of all explanatory variables for prediction using the random forests permutation-based score of importance, and finalized with a forward step-wise strategy for variable introduction.

Here we utilize VSURF to select unique sets of variables for each crop (barley, maize, wheat, teff). Given the complexity of these permutations for each crop (100s of millions) we utilize 15 cores of the GWU supercomputer Colonial One. Each run took between 10-24 hours.

### **Regression**

Compared to cross-sectional approaches, panel analysis can substantially increase the degree of observed variance over both space and time. For instance, in our current application, the integration of EAS (with  $n = 1752$ ) over the 2010-2015 sample period ( $t = 6$ ), results in a substantially larger data ( $N \sim 6900$ ), allowing for a much greater flexibility and degrees of freedom in the modelling of the issue of interest. Here we include a few variables of interest to the reader including Zonal and annual dummies, and a lagged value of output per hectare. Here we estimate a series of random effect models of the following form 1:

(1)

$$Y_{itc} = \alpha_i + \sum_{k=1}^K \beta_k X_{it} + \mu_{it} + \epsilon_{it}$$

where  $Y_{itc}$  is the output per hectare (OPH) for crop  $c$  for sub-kebele  $i$  for year  $t$ ,  $\alpha$  is the average output per hectare for sample population (country),  $\sum_{k=1}^K \beta_k X_{it}$  is the sum of all  $K$  coefficient estimates  $\beta_k$  for all variables  $X$ ,  $\mu_i$  is the sub-kebele specific random effect that measures the difference between the average OPH at sub-kebele  $i$  and the average OPH in the entire country. The term  $\epsilon_{it}$  is the individual-specific effect, which can be described as the deviation of the  $i$ th OPH from the average for that year.

Here we present preliminary results from panel regressions on output per hectare for Ethiopia's major crops. These variables reflect the initial selection from the VSURF algorithm.

## Predicting Crop Losses

Another critical emphasis of this project is to see if remotely sensed data on greenness, precipitation, or evapotranspiration and the zone code can be used in conjunction with household surveys to predict the location and/or extent of damages before the next years AgSS survey results are released. In this vein we aim to utilize advances from machine learning to accurately predict substantial crop damages for all major crops. We define our dependent variable, ‘substantial’ losses, as farmer reported losses over 30%.

### Estimation Strategy

#### Random Forests Description

Random forests (RF) are a flexible (used for classification and regression) ensemble learning method that aggregates across multiple classification or regression trees each based on a randomized subset of variables. Each tree is formed through hierarchical splitting method where by each binary split (e.g. NDVI < 0.1) is identified as the most informative, as defined by the greatest reduction in Gini. This ensemble method improves performance by finding the mode or median of ‘weak learners’. This process in turn allows RFs to avoid over-fitting and therefore perform consistently out-of-sample and with noisy data sets.

#### Training vs Testing Data

Before we run any models we break the data into two groups, with 75% of all sub-kebele units being retained for the training/tuning/cross-validation of the models, and 25% of these held aside to provide independent measures of out-of-sample performance.

#### Training Data: Cross Validation & Tuning

To enhance the performance of the models out-of-sample, we undertake two procedures on the training data; tuning and cross validation. Tuning in this case is limited to choosing the number of features (variables) randomly selected for use in each tree. We do a grid search independently for each crop finding the parameter tuning that provides the best performance. To ensure that these parameter tunings work well, we utilize k-fold cross validation whereby the training data is split further into 3 sets of sub-kebele data, the model is trained on two thirds of the data and its performance, using the Kappa coefficient, is estimated using the omitted third sub-sample. Performance for each parameter value is therefore evaluated out-of-sample three times and averaged. The optimal parameter then is that with the highest Kappa value and is specific for each crop model. Kappa<sup>2</sup> is a performance indicator similar to the percentage of observations that are correctly classified but that controls for the fact that categories with large numbers of observations are easier to guess by random chance.

#### Testing Data: Out-of-sample Performance

Independent of the training data we initially withheld 25% of sub-kebele units to provide ‘testing data’ for accurate measures of truly out-of-sample performance for our tuned RF models. These tests should therefore be good indicators of performance for sub-kebeles excluded from the AgSS survey. To test accuracy we make predictions of the damage category on this testing data and compare them to each observations observed damage category. We provide three major performance indicators, a) overall accuracy, 2) Kappa coefficient, and 3) the recall rate<sup>3</sup>.

<sup>2</sup>see: [https://en.m.wikipedia.org/wiki/Cohen%27s\\_kappa](https://en.m.wikipedia.org/wiki/Cohen%27s_kappa)

<sup>3</sup>the percentage of sub-kebeles with ‘substantial damage’ that are correctly classified by our models.

## Geographically Weighted Regression

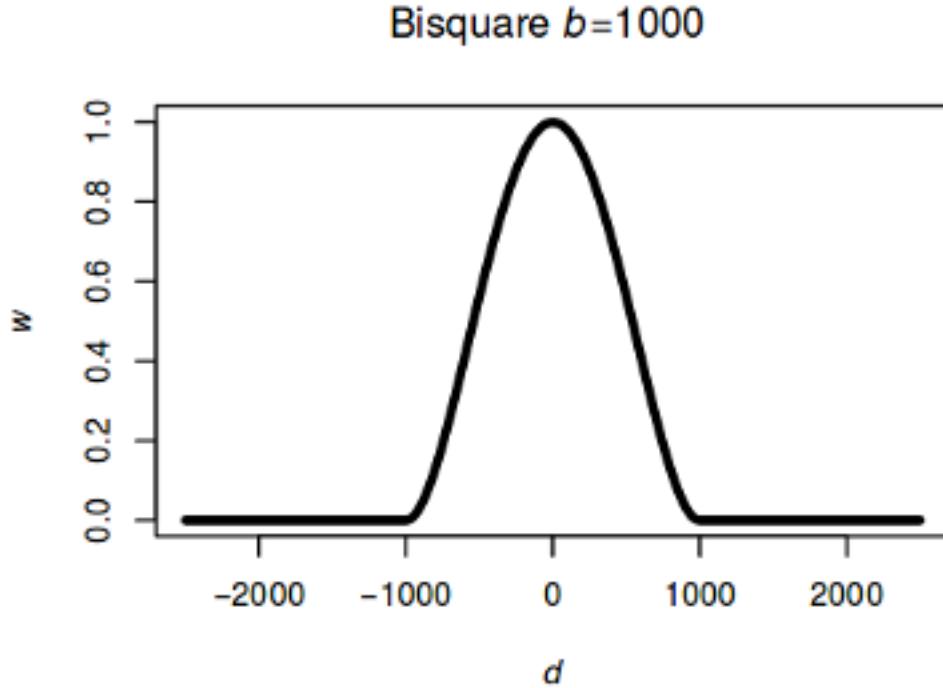
Another way to explore the impacts of policy efforts is to look at how their impacts vary across space. We can do this by estimating how the relationships between a dependent and independent variables change in different areas. One popular method is using Geographically Weighted Regression (GWR) (Brunsdon, Fotheringham, and Charlton 1996, Brunsdon, Fotheringham, and Charlton (1998)). A simple GWR implementation is as follows 2:

(2)

$$Y_i = \alpha_i + \sum_{k=1}^m \beta_{ik} X_{ik} + \epsilon_i$$

where  $Y_i$  is the dependent variable at location  $i$ ,  $\alpha_i$  is the local intercept value,  $\beta_{ik}$  is the local estimate of a coefficient for the  $k$ th independent variable at location  $i$  times  $\beta_{ik}$ , which is the local coefficient estimate for the same location and independent variable, where  $m$  is number of independent variables, and  $\epsilon_i$  is the random error component for location  $i$ .

These estimates are weighted by some geographic weighting function, whereby nearby locations are given greater weight in estimating local coefficient values than those farther away. In this case we use a bi-square adaptive kernel, where closer observations are given weights near one and further observations given less weight as follows in the Figure 5. We use an ‘adaptive’ kernel, indicating that the distance used to assigned weights can vary across the surface, but the number of observations used is fixed at some number. We use the adaptive kernel since some of our samples are geographically isolated, and therefore would have problems with local degrees of freedom.



of bi-square weighting kernel

Figure 5: Example

GWR can suffer from a number of problems in estimation, including local outliers and local multicollinearity. We address the prior with a robust implementation of GWR (Gollini et al. 2013), that removes local outliers by examining the studentized residuals from an initial GWR estimation.

Key take aways for GWR: \* A new regression is estimated for every location \* Each location has it's own coefficient estimates, standard errors, and R^2 \* Local outliers can be addressed

# Results

## Regression Results

Table 1: Panel RE regression on output per hectare for maize in quintals

Panel RE regression

	Dependent variable: MAIZEOPH_W			
	Full (1)	Reduced (2)	Final (3)	No AgSS (4)
MAIZEOPH_W, 1)	0.3262*** (0.0101)	0.3227*** (0.0101)	0.3265*** (0.0100)	0.3653*** (0.0103)
G_mn	0.0029** (0.0014)	0.0021 (0.0013)	0.0025* (0.0013)	0.0029** (0.0013)
G_AUC_leading	0.000004 (0.000004)			
G_AUC_Qnt	0.000003 (0.000002)			
A_mn	-0.0053*** (0.0017)	-0.0045*** (0.0016)	-0.0034*** (0.0012)	-0.0049*** (0.0013)
A_max	-0.00003 (0.0005)			
A_AUC	0.0000* (0.0000)	0.0000 (0.0000)		
A_Qnt	0.0018*** (0.0006)	0.0020*** (0.0005)	0.0027*** (0.0004)	0.0048*** (0.0004)
A_AUC_Qnt	-0.0000 (0.0000)			
A_max_Qnt	-0.0010** (0.0004)	-0.0010*** (0.0004)	-0.0007** (0.0003)	-0.0014*** (0.0003)
T_G_Qnt	0.0013* (0.0007)	0.0012* (0.0006)		
PPT_A_max	0.0010 (0.0083)	0.0001 (0.0078)		
PPT_G_AUC_Qnt	0.0020** (0.0010)	0.0023*** (0.0007)	0.0024*** (0.0007)	0.0027*** (0.0007)
PPT_G_mx_Qnt	-0.0204** (0.0094)	-0.0183** (0.0089)	-0.0189** (0.0083)	-0.0309*** (0.0086)
PPT_T_G_Qnt	0.0201 (0.0222)			
PET_A_min	0.0108*** (0.0032)	0.0125*** (0.0032)	0.0118*** (0.0031)	0.0169*** (0.0032)
PET_G_AUC_diff_mn	0.0006*** (0.0001)	0.0005*** (0.0001)	0.0006*** (0.0001)	0.0003** (0.0001)
MAIZEAREA	0.1171*** (0.0440)	0.0545 (0.0432)		
MAIZESEED1AREA	-0.0195 (0.1486)			
Fert_Amt_Per_Area		0.0015***	0.0015***	

		(0.0004)	(0.0004)
MAIZEIMSEED	0.0216*** (0.0047)	0.0186*** (0.0031)	0.0191*** (0.0031)
MAIZEDAMAGEAREA_P	-11.1709*** (0.5803)	-11.1543*** (0.5715)	-11.0321*** (0.5673)
MAIZEFERT_CHEMICAL_AREA	-0.2302** (0.0970)		
MAIZEFERT_CHEMICAL_AMT	0.0002 (0.0001)		
MAIZEEXTAREA	0.2161** (0.0936)	0.1061 (0.0766)	0.1513** (0.0668)
Y_COORD	-0.000001 (0.000003)	0.000001 (0.000003)	
X_COORD	0.00001*** (0.000003)	0.00001*** (0.000003)	0.00001*** (0.000003) 0.000004 (0.000003)
elevation	-0.0019*** (0.0003)		
bs(elevation, 3)1		15.4945*** (3.8604)	15.5929*** (3.7878) 23.8067*** (3.9425)
bs(elevation, 3)2		4.6195*** (1.5420)	4.5506*** (1.5263) 7.7403*** (1.5913)
bs(elevation, 3)3		4.3731 (2.6649)	4.3404* (2.6305) 8.6528*** (2.7460)
c(Year)	0.7367*** (0.1042)	0.7053*** (0.1031)	0.6262*** (0.0813) 0.5059*** (0.0838)
Constant	2.0693 (5.0900)	-9.9053* (5.5259)	-7.2621** (2.9705) -17.1149*** (3.0796)

Zone Dummy	Yes	Yes	Yes	Yes
Observations	8,241	8,241	8,241	8,241
R2	0.4630	0.4675	0.4696	0.4211
Adjusted R2	0.4576	0.4624	0.4648	0.4163
F Statistic	84.7407***	91.8726***	99.0421***	86.1521***

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Looking at the results from Table 1 with regression results from Maize OPH (quintals per hectare) we can see the results from four model outputs. The first, “Full”, presents the results based on the model selection from the VSURF algorithm and includes zonal fixed effects (FE). The second, “Reduced”, narrows the variable selection while including critical non-linearity for variables such as elevation, and adds a variable *Fert\_Amt\_Per\_Area* which is the amount of chemical fertilizer applied per unit area. Non-linearity is controlled for using basis-splines which outperform polynomial representations in terms of stability and forecasting. The third model, “Final”, represents the final selection of significant variables for estimation. The last model, “No AgSS”, omits all AgSS survey variables to demonstrate the ability of the model to predict yields on the basis of remotely sensed data and last years yields.

Looking at the final model we can see an impressive adjusted R^2 of 0.46. Residuals are also relatively small with 1st and 3rd quintiles spanning -7.66 and 6.56 quintals. These numbers are relatively impressive considering the highly-localized nature of these estimates (at the sub-kebele level). Looking more closely we can observe some important features.

First, remotely sensed variables, while complex, have a few tangible interpretations. For instance, *G\_mn* is positive and significant ( $p<0.0503$ ) indicating that yields are higher if the mean greenness (NDVI) level during the Meher growing season is higher. Similarly, increases in the 90th percentile of greenness for the

entire year ( $A_{Qnt}$ ), corresponds to statistically significant higher yields ( $p<1.66e-12$ ). Interestingly, the variable  $A_{mn}$ , which is the mean greenness level (NDVI) for the entire year, is negative and significant ( $p<0.00498$ ). Already controlling for conditions during the Meher season, this might be an indication of two issues, a) a higher  $A_{mn}$  might indicate the failure to properly till between seasons, or b) complications of management in dual-crop areas, where planting occurs for both the minor (Belg) and major (Meher) rainy season.

Looking at precipitation variables we can see that  $PPT_G_AUC_Qnt$  yields are higher ( $p<0.000716$ ) in sub-kebeles where the 90th percentile of total rainfall during the Meher season is higher. The opposite is true ( $p<0.0223$ ) for areas with higher 90 percentile maximum rainfalls ( $PPT_G_mx_Qnt$ ) likely reflecting the impact of flooding or excess rain events. Additionally, areas with higher minimum potential evaporation rates ( $PET_A_min$ : higher min temps, more sun) receive better yields ( $p<0.000137$ ) while controlling for precip and greenness, and areas with above average total PET ( $PET_G_AUC_diff_mn$ ) also receive yield benefits ( $p<7.83e-06$ ).

The AgSS data also provide some insights. We see scant evidence for yield improvements based on planted area alone ( $MAIZEAREA$   $p<0.207$ ), but we do see the positive effects of fertilization rates ( $Fert_Amt_Per_Area$   $p<0.000357$ ), the use of improved seeds ( $MAIZEIMSEED$   $p<6.31e-10$ ), and the area of land in the sub-kebele that has utilized extension agents ( $MAIZEEXTAREA$   $p<0.0234$ ). Conversely, we see the substantial impact of the percent of reported crop damage ( $MAIZEDAMAGEAREA_P$ ), with every additional one percent of land damaged corresponding to a -11 unit loss in OPH ( $p<2.21e-82$ ).

Table 2: Panel RE regression on output per hectare for wheat in quintals

Panel RE regression

	Dependent variable:			
	WHEATOPH_W			
	Full (1)	Reduced (2)	Final (3)	No AgSS (4)
WHEATOPH_W, 1)	0.2803*** (0.0123)	0.2797*** (0.0123)	0.2765*** (0.0122)	0.2952*** (0.0124)
G_mx	0.0012*** (0.0004)	0.0012*** (0.0004)	0.0012*** (0.0004)	0.0027*** (0.0004)
A_max_Qnt	-0.0010** (0.0004)	-0.0011*** (0.0004)	-0.0011*** (0.0004)	-0.0014*** (0.0004)
T_G_Qnt	0.0014*** (0.0005)	0.0015*** (0.0005)	0.0015*** (0.0005)	0.0006 (0.0005)
PPT_G_Qnt	-0.0544* (0.0279)			
PPT_G_sd	0.2214*** (0.0766)	0.0906*** (0.0248)		
PPT_G_mx_Qnt	-0.0210** (0.0096)	-0.0124 (0.0092)		
PPT_G_AUC_Qnt	0.0020* (0.0011)			
PPT_T_G_Qnt	-0.0943*** (0.0244)	-0.0720*** (0.0192)	-0.0540*** (0.0120)	-0.0543*** (0.0123)
PET_G_mn	0.0075** (0.0038)			
PET_G_sd	-0.0384*** (0.0106)			
WHEATAREA	0.0619	0.1389***	0.1441***	

	(0.0733)	(0.0242)	(0.0239)
WHEATNIMSEED	0.0005*		
	(0.0003)		
WHEATNIMSEED_P	0.0008	0.0011	0.0011
	(0.0008)	(0.0008)	(0.0008)
WHEATDAMAGEAREA_P	-9.9673***	-9.8878***	-9.9924***
	(0.6623)	(0.6607)	(0.6605)
WHEATFERT_CHEMICAL_AMT	0.0007***		
	(0.0002)		
WHEATFERT_CHEMICAL_AREA	-0.0800		
	(0.0728)		
WHEATFERT_CHEMICAL_AMT_P	-0.0001		
	(0.0001)		
Fert_Amt_Per_Area		0.0007	
		(0.0006)	
WHEATFERT_CHEMICAL_AREA_P	1.8758***	1.6178***	1.6394***
	(0.3841)	(0.3345)	(0.3329)
X_COORD	0.00001***	0.00001***	0.00001***
	(0.000004)	(0.000004)	(0.000003)
Y_COORD	0.000004	-0.000000	
	(0.000003)	(0.000003)	
dist_rcap	0.000002		
	(0.000004)		
elevation	0.0016***		
	(0.0003)		
bs(elevation, 3)1		8.9310***	9.2846***
		(2.4111)	(2.3904)
bs(elevation, 3)2		4.8475***	4.8974***
		(1.2356)	(1.2294)
bs(elevation, 3)3		6.8167***	6.8253***
		(1.9207)	(1.9176)
c(Year)	1.0793***	1.2141***	1.0587***
	(0.0937)	(0.0807)	(0.0673)
Constant	-11.7601*	-6.9866	-6.5465
	(6.7969)	(6.7235)	(4.5185)
13.0284***			
(2.4531)			
8.7619***			
(1.2135)			
8.5772***			
(1.9746)			
1.0202***			
(0.0686)			
-11.0476**			
(4.6501)			

Zone Dummy Yes Yes Yes Yes

Observations	5,660	5,660	5,660	5,660
R2	0.4042	0.4044	0.4024	0.3651
Adjusted R2	0.3962	0.3970	0.3955	0.3582
F Statistic	50.5062***	54.9956***	57.9550***	52.7664***

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 3: Panel RE regression on output per hectare for barley in quintals

#### Panel RE regression

Dependent variable:

	BARLEYOPH_W			
	Full (1)	Reduced (2)	Final (3)	No AgSS (4)

BARLEYOPH_W, 1)	0.2247*** (0.0122)	0.2255*** (0.0122)	0.2258*** (0.0122)	0.2394*** (0.0124)
G_mx	0.0007* (0.0004)	0.0008* (0.0004)	0.0008** (0.0004)	0.0011*** (0.0004)
G_Qnt	0.0020*** (0.0004)	0.0020*** (0.0004)	0.0019*** (0.0004)	0.0026*** (0.0004)
A_AUC_Qnt	-0.0000 (0.0000)	-0.0000 (0.0000)		
A_max_Qnt	-0.0015*** (0.0003)	-0.0015*** (0.0003)	-0.0015*** (0.0003)	-0.0020*** (0.0003)
PPT_A_max	-0.0008 (0.0140)	-0.0004 (0.0140)		
PPT_G_AUC	-0.0038 (0.0024)	-0.0040* (0.0024)	-0.0041* (0.0023)	-0.0051** (0.0024)
PPT_A_mn	0.1328 (0.0930)	0.1274 (0.0930)	0.1315 (0.0916)	0.1923** (0.0936)
PPT_G_sd	-0.2603* (0.1413)	-0.2503* (0.1413)	-0.2585* (0.1406)	-0.3184** (0.1437)
PPT_A_sd	0.3365** (0.1570)	0.3216** (0.1567)	0.3300** (0.1514)	0.3940** (0.1547)
PPT_G_AUC_leading	-0.0005 (0.0004)			
PPT_G_AUC_Qnt	0.0028*** (0.0010)	0.0027*** (0.0010)	0.0029*** (0.0010)	0.0018* (0.0010)
PPT_G_mx_Qnt	-0.0228*** (0.0078)	-0.0205** (0.0080)	-0.0197** (0.0077)	-0.0206*** (0.0078)
PPT_T_G_Qnt	-0.0717*** (0.0199)	-0.0725*** (0.0199)	-0.0730*** (0.0192)	-0.0653*** (0.0194)
PET_G_mx	0.0009 (0.0023)	0.0009 (0.0023)		
BARLEYFERT_CHEMICAL_AMT	0.0003 (0.0004)			
Fert_Amt_Per_Area		0.00004 (0.0006)		
BARLEYDAMAGEAREA_P	-7.4954*** (0.4995)	-7.4732*** (0.4999)	-7.4443*** (0.4967)	
BARLEYFERT_CHEMICAL_AREA	0.2160*** (0.0638)	0.2413*** (0.0514)	0.2322*** (0.0504)	
BARLEYFERT_NATURAL_AREA	-0.0454 (0.1131)	-0.0139 (0.1157)	-0.0218 (0.1154)	
Y_COORD	-0.000004 (0.000003)	-0.000004 (0.000003)	-0.000004 (0.000003)	-0.00001* (0.000003)
X_COORD	0.00001*** (0.000003)	0.00001*** (0.000003)	0.00001*** (0.000003)	0.00001*** (0.000003)
dist_rcap	0.000004 (0.000003)	0.000004 (0.000003)		
elevation	0.0030*** (0.0003)			
bs(elevation, 3)1		1.8205 (2.0948)	1.6627 (2.0841)	2.2304 (2.1006)
bs(elevation, 3)2		6.2227*** (0.9847)	6.1114*** (0.9739)	7.6853*** (0.9756)
bs(elevation, 3)3		6.4290***	6.2997***	7.9813***

		(1.5308)	(1.5200)	(1.5262)
c(Year)	0.5863***	0.5953***	0.5936***	0.4895***
	(0.0830)	(0.0830)	(0.0828)	(0.0843)
Constant	2.5650	6.4145	6.9160	5.1132
	(4.7692)	(4.8473)	(4.8024)	(4.8541)
Zone Dummy	Yes	Yes	Yes	Yes
Observations	5,864	5,864	5,864	5,864
R2	0.3992	0.3998	0.4002	0.3788
Adjusted R2	0.3915	0.3919	0.3929	0.3715
F Statistic	51.2850***	50.7198***	54.4388***	51.9587***

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 4: Panel RE regression on output per hectare for teff in quintals

Panel RE regression

Dependent variable:				
	TEFFOPH_W			
	Full (1)	Reduced (2)	Final (3)	No AgSS (4)
TEFFOPH_W, 1)	0.2873*** (0.0109)	0.2873*** (0.0109)	0.2882*** (0.0108)	0.3113*** (0.0110)
G_mx_Qnt	-0.0001 (0.0036)			
G_AUC_diff_mn	0.000004 (0.00001)	0.000003 (0.00001)		
G_AUC_Qnt	0.00001* (0.00001)	0.00001*** (0.000003)	0.00001*** (0.000003)	0.00001** (0.000003)
A_sd	0.0020*** (0.0004)	0.0022*** (0.0003)	0.0023*** (0.0003)	0.0029*** (0.0003)
A_max_Qnt	0.0002 (0.0039)			
T_G_Qnt	0.0001 (0.0004)			
PPT_G_AUC	-0.0004 (0.0006)			
PPT_G_AUC_trailing	-0.0002 (0.0003)			
PPT_G_sd	0.0273 (0.0207)	0.0190 (0.0168)		
PPT_G_mx_Qnt	-0.0118** (0.0060)	-0.0079 (0.0058)		
PPT_G_AUC_Qnt	0.0019** (0.0008)	0.0015** (0.0006)	0.0013** (0.0006)	0.0018*** (0.0006)
PET_A_sd	-0.0151** (0.0063)	-0.0155** (0.0062)	-0.0167*** (0.0061)	-0.0223*** (0.0062)
PPT_T_G_Qnt	-0.0336** (0.0149)	-0.0331** (0.0147)	-0.0328*** (0.0126)	-0.0299** (0.0128)
TEFFDAMAGEAREA_P	-7.2108***	-7.0764***	-7.0871***	

	(0.4368)	(0.4351)	(0.4349)
TEFFFERT_CHEMICAL_AMT	0.0001 (0.0001)	0.0003** (0.0001)	0.0003** (0.0001)
Fert_Amt_Per_Area		-0.0010** (0.0005)	-0.0010** (0.0005)
TEFFAREA	0.1034*** (0.0148)	0.0807*** (0.0163)	0.0805*** (0.0163)
Y_COORD	-0.000004* (0.000002)	-0.000003* (0.000002)	-0.000003* (0.000002) -0.000005** (0.000002)
X_COORD	0.00001*** (0.000002)	0.00001*** (0.000002)	0.00001*** (0.000002) 0.00001*** (0.000002)
dist_pp50k	0.000001 (0.000003)		
dist_rcap	0.00001** (0.000002)	0.00001*** (0.000002)	0.00001*** (0.000002) 0.00001** (0.000002)
elevation	0.0019*** (0.0002)		
bs(elevation, 3)1		-6.2332*** (2.0305)	-6.3313*** (2.0264) -5.1210** (2.0549)
bs(elevation, 3)2		6.8509*** (0.8863)	6.8793*** (0.8823) 9.3317*** (0.8802)
bs(elevation, 3)3		-2.7897* (1.6085)	-2.9933* (1.6004) -3.4363** (1.6195)
c(Year)	0.6047*** (0.0597)	0.5814*** (0.0587)	0.5414*** (0.0478) 0.4335*** (0.0484)
Constant	0.1557 (3.3280)	4.5775 (3.2890)	5.0360 (3.2636) 5.6933* (3.3054)

Zone Dummy	Yes	Yes	Yes	Yes
Observations	7,373	7,373	7,373	7,373
R2	0.3039	0.3094	0.3100	0.2819
Adjusted R2	0.2965	0.3024	0.3033	0.2753
F Statistic	41.3527***	44.1837***	46.1929***	42.8010***

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 5: Panel RE regression on output per hectare for sorghum in quintals

#### Panel RE regression

Dependent variable: SORGHUMOPH_W				
	Full (1)	Reduced (2)	Final (3)	No AgSS (4)
SORGHUMOPH_W, 1	0.2897*** (0.0120)	0.2916*** (0.0120)	0.2910*** (0.0120)	0.2933*** (0.0123)
A_Qnt	0.0023*** (0.0003)	0.0023*** (0.0003)	0.0020*** (0.0002)	0.0027*** (0.0002)
G_mn	-0.0004 (0.0003)	-0.0004 (0.0003)		
PPT_T_G_Qnt	0.0102	0.0078		

	(0.0123)	(0.0125)		
PET_A_min	-0.0011 (0.0032)			
SORGHUMDAMAGEAREA_P	-9.8842*** (0.6497)	-9.8835*** (0.6481)	-9.8568*** (0.6479)	
SORGHUMFERT_CHEMICAL_AMT	-0.00002 (0.0002)			
Fert_Amt_Per_Area		0.0002 (0.0009)		
Y_COORD	0.00001*** (0.000003)	0.00001*** (0.000003)	0.00001*** (0.000003)	0.00001*** (0.000003)
X_COORD	0.00001*** (0.000003)	0.00001*** (0.000003)	0.00001*** (0.000003)	0.00001*** (0.000003)
dist_rcap	0.00001** (0.000004)	0.00001** (0.000004)	0.00001** (0.000004)	0.000004 (0.000004)
elevation	-0.0011*** (0.0004)			
bs(elevation, 3)1		3.8736 (4.1331)	4.1269 (4.0369)	3.7254 (4.1162)
bs(elevation, 3)2		-1.5559 (1.7381)	-1.6717 (1.6432)	-0.9546 (1.6749)
bs(elevation, 3)3		0.5675 (2.9958)	0.7859 (2.8937)	0.6827 (2.9505)
c(Year)	0.8758*** (0.0790)	0.8647*** (0.0732)	0.8510*** (0.0724)	0.7102*** (0.0731)
factor(REGIONCODE)2	0.7158 (3.2708)			
factor(REGIONCODE)3	-0.2054 (1.6132)			
factor(REGIONCODE)4	2.8698 (3.4585)			
factor(REGIONCODE)6	2.9337 (2.1518)			
factor(REGIONCODE)7	1.6541 (2.6375)			
Constant	-15.1083*** (5.5267)	-18.1295*** (6.0125)	-18.0528*** (6.0107)	-21.9174*** (6.1235)

Zone Dummy	Yes	Yes	Yes	Yes
Observations	6,096	6,096	6,101	6,101
R2	0.3363	0.3386	0.3379	0.3121
Adjusted R2	0.3292	0.3313	0.3310	0.3050
F Statistic	47.0116***	46.7556***	48.9108***	44.1786***

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Predicting damages

### Maize Damages

We can evaluate accuracy with a number of metrics on our independent testing data set described in the “Testing Data: Out-of-sample Performance” section above. The first, the confusion matrix compares predicted

values against the observed (reference). We also report the a) overall accuracy, 2) Kappa coefficient, and 3) the recall rate<sup>4</sup> as seen in table 6 below:

Table 6: Confusion matrix for reported damages > 30% for Maize

Table 1: FALSE = Crop damages greater than 30%

	FALSE	TRUE
FALSE	1516	121
TRUE	220	4791

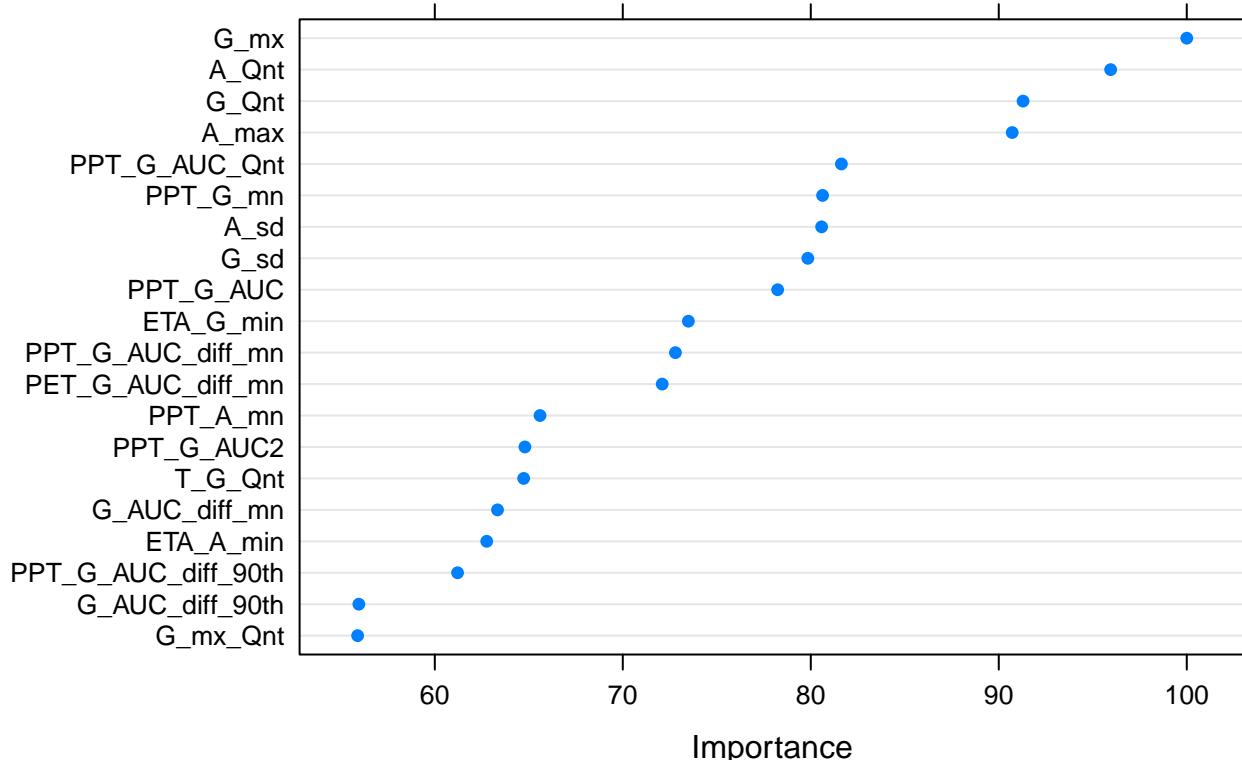
Table 7: Performance metrics for reported damages > 30% for Maize

	Accuracy	Kappa	Recall
Value	0.95	0.86	0.87

We see impressive identification of reported maize crop losses with a recall rate of 0.87, indicating that 87% of all substantial loss cases were predicted in our completely independent *testing* data. In all, 1516 out of 1736 reported cases were predicted, and only 121 cases were falsely predicted as crop failure areas.

We can also examine the relative importance of each variable to each model, as well as its estimated non-linear relationship with the dependent variable through partial dependency plots. First, let's examine the relative importance of each variable in the random forest through the role it played in Gini coefficient loss.

Figure 6: Variable importance plot for maize damage prediction



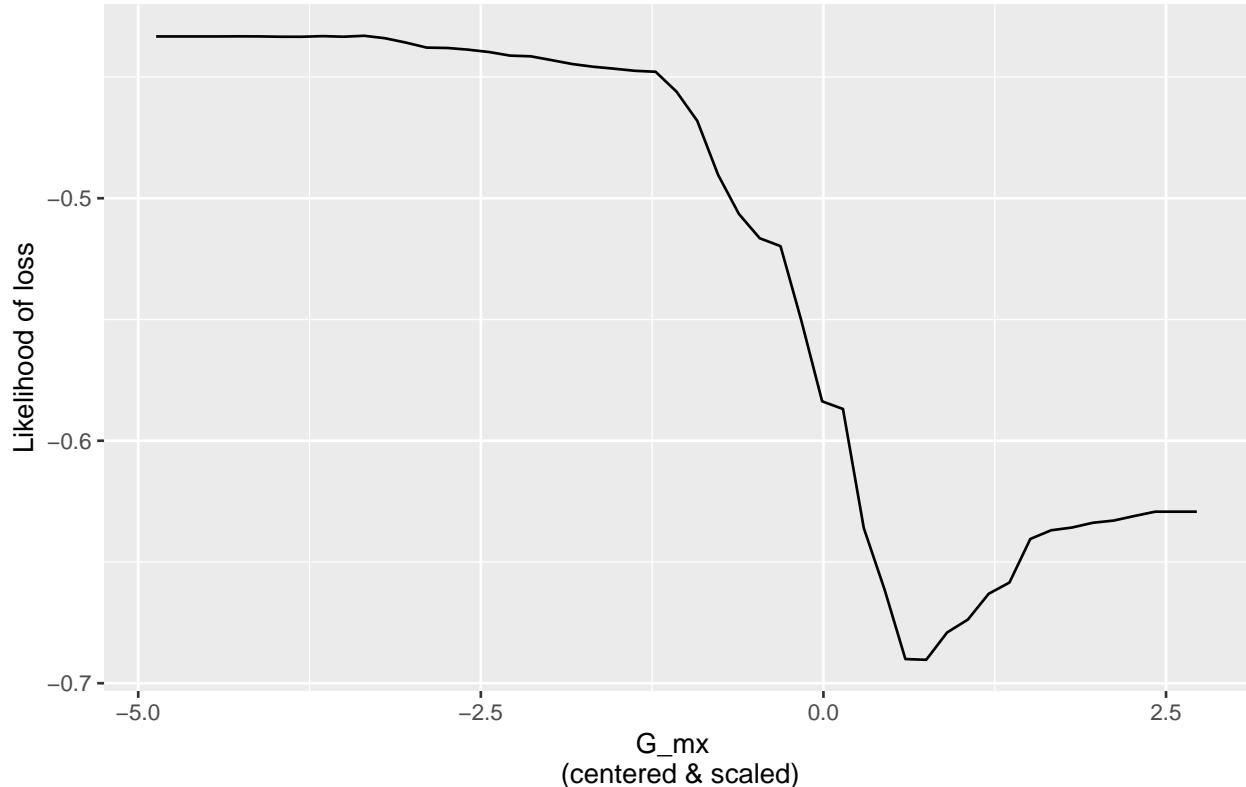
From the figure above we can see that G\_mx has the highest relative importance in the random forest model.

<sup>4</sup>the percentage of sub-kebeles with ‘substantial damage’ that are correctly classified by our models.

$G_{mx}$  is measured as the maximum NDVI value for the growing season. As such the highest NDVI value likely reflects a measure of plant health, with higher  $G_{mx}$  values corresponding to strong plant health at some point in the growing season. We can also see that the most important variables are largely comprised of variables summarizing NDVI, with the exception of  $PPT\_G\_{mn}$  which is the mean value of precipitation for each 10 day period across the growing season.

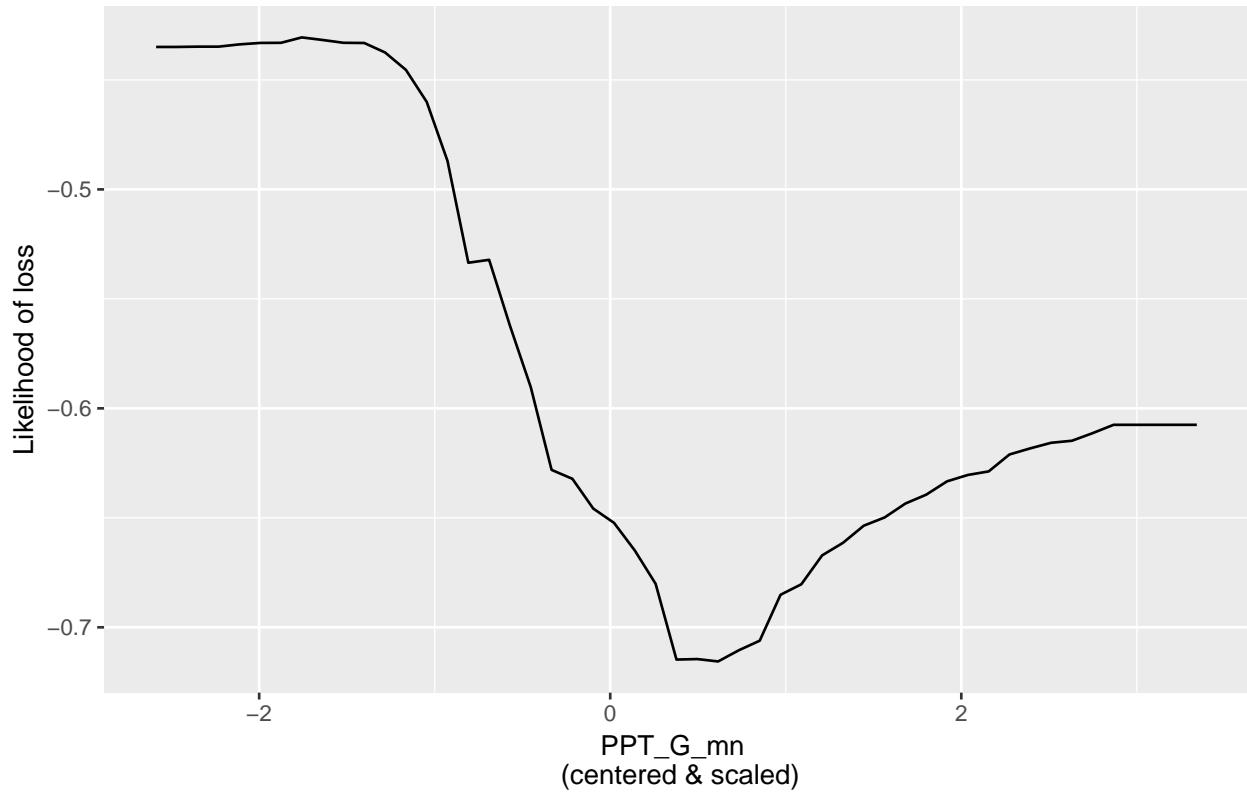
We can now look at the partial dependence plots to help understand the non-linear response of reported damages to variables of interest, such as  $G_{mx}$ .

Figure 7: Partial dependency plot for influence of  $G_{mx}$  on maize damage prediction



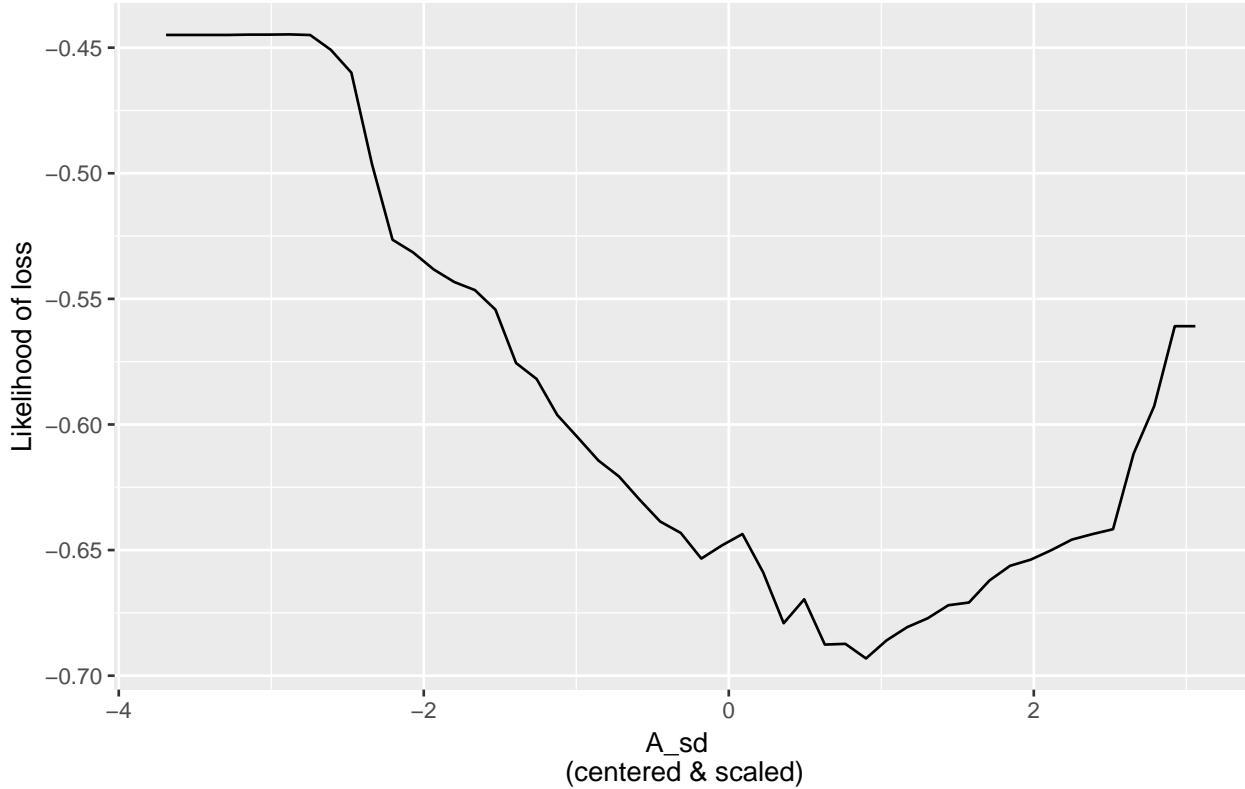
In the figure above, we can see how the likelihood of reporting maize losses decreases rapidly after one standard deviation below the mean value of  $G_{mx}$ , with the likelihood of losses increasing at the upper end of the distribution. These losses at upper end of the distribution might reflect excess rainfall, or late season losses from lodging or frost. Ideally additional variables would capture the effects of these late season dynamics in ways that maximum NDVI cannot. Similarly, we can look at the partial dependence plot for other variables of interest such as  $PPT\_G\_{mn}$ .

Figure 8: Partial dependency plot for influence of  $PPT\_G\_{mn}$  on maize damage prediction



$PPT\_G\_mn$  tracks the effects of the average 10-day precipitation across the growing season. We can see that low levels of precipitation have the highest likelihood of crop damage. The likelihood of failure declines rapidly at about one standard deviation below the mean, and reaches its lowest likelihood around  $+0.5$  sd above the mean precipitation rate. Past above  $+1$  sd higher rainfall rates again increase the likelihood of damage, but at a slower rate. We can see how variance in NDVI corresponds to crop failures with variables such as  $A\_sd$ .

Figure 9: Partial dependency plot for influence of  $A\_sd$  on maize damage prediction



$A_{sd}$  is a measure of variability in NDVI across the entire year (including both the Meher and Belg seasons). We see a near linear decrease in the likelihood of damage starting from 1.5 sd below the mean, with a minima near one sd above. The high likelihood of failure corresponding with very low values of NDVI variance likely relates to early season trouble including a failure to germinate, or early damage leading to crop failure. These early catastrophic losses would have little change in NDVI across the year and therefore very low values of  $A_{sd}$ . After the minima, we see a minima in damages around +1 sd. This is likely related to the fact that a plot of healthy plants that is harvested and tilled, will likely have relatively high variance (especially in dual-cropping Meher and Belg areas). We can see around +2 sd that extremely high variance again corresponds to increased likelihood of losses, as dramatic shifts in plant health would likely correspond to.

## Wheat Damages

The confusion matrix compares predicted values against the observed (reference). We also report the a) overall accuracy, 2) Kappa coefficient, and 3) the recall rate<sup>5</sup> as seen in table 8 below:

Table 8: Confusion matrix for reported damages > 30% for Wheat

Table 3: FALSE = Crop damages greater than 30%

	FALSE	TRUE
FALSE	1028	16
TRUE	131	3457

Table 9: Performance metrics for reported damages > 30% for Wheat

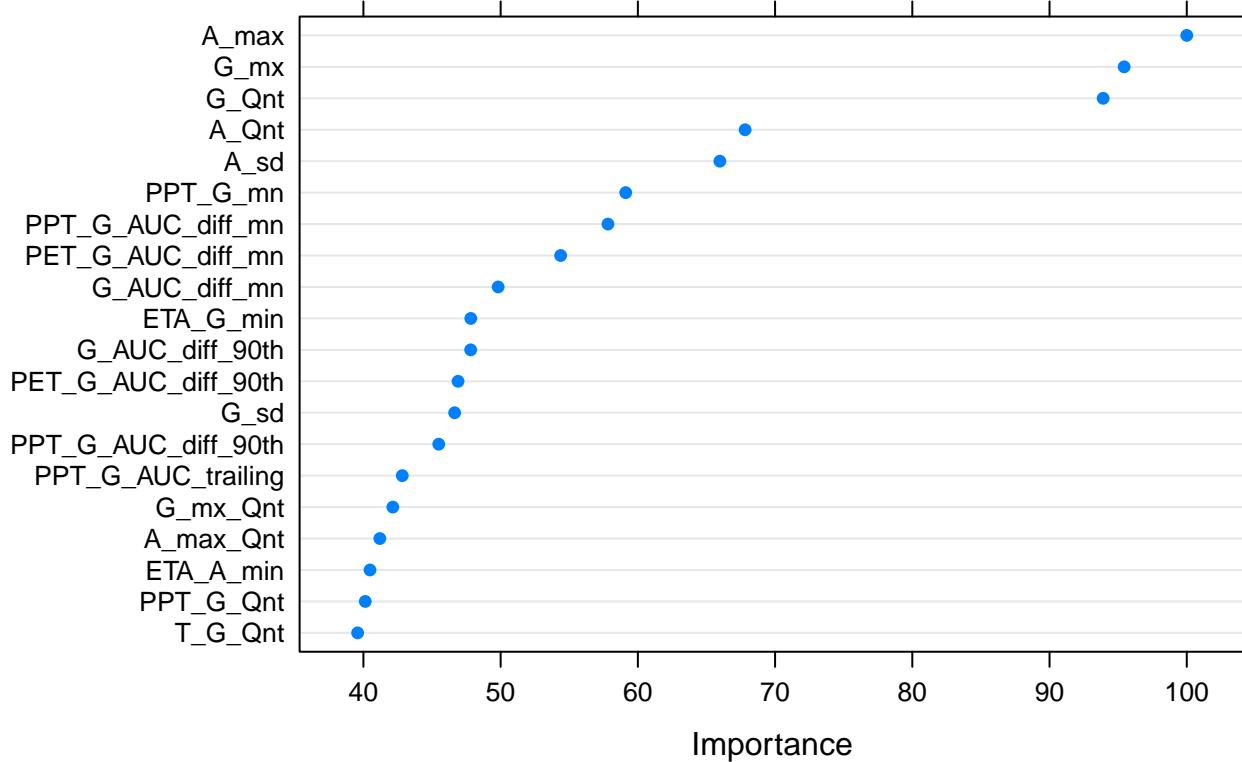
<sup>5</sup>the percentage of sub-kebeles with ‘substantial damage’ that are correctly classified by our models.

	Accuracy	Kappa	Recall
Value	0.97	0.91	0.89

For the case of wheat, we see impressive identification of reported wheat crop losses with a recall rate of 0.89, indicating that 89% of all substantial loss cases were predicted in our completely independent *testing* data. In all, 1028 out of 1159 reported cases were predicted, and only 16 cases were falsely predicted as crop failure areas.

We can also examine the relative importance of each variable to each model, as well as its estimated non-linear relationship with the dependent variable through partial dependency plots. First, let's examine the relative importance of each variable in the random forest through the role it played in Gini coefficient loss.

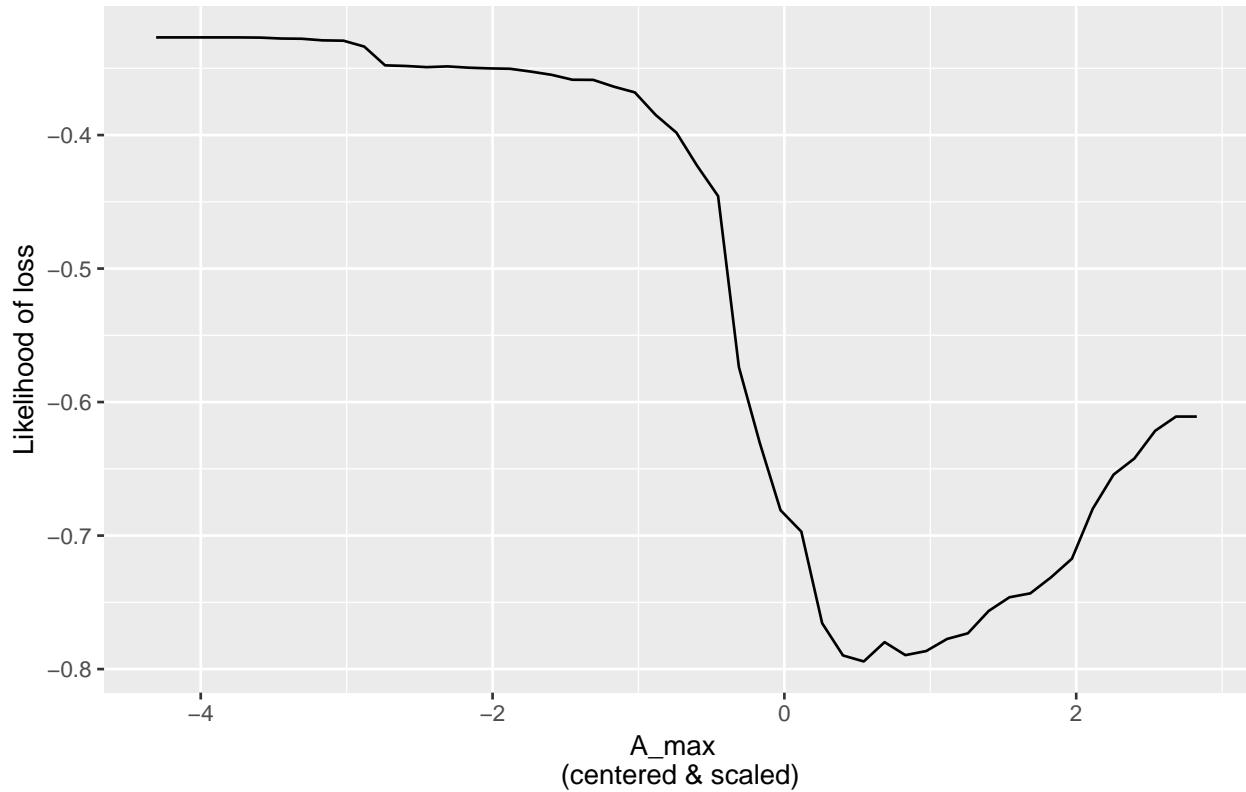
Figure 10: Variable importance plot for maize damage prediction



From the figure above we can see that **A\_max** has the highest relative importance in the random forest model. **A\_max** is measured as the maximum NDVI value for the growing season. As such the highest NDVI value likely reflects a measure of plant health, with higher **A\_max** values corresponding to strong plant health at some point in the growing season. We can also see that the most important variables are largely comprised of variables summarizing NDVI, with the exception of **PPT\_G\_mn** which is the mean value of precipitation for each 10 day period across the growing season.

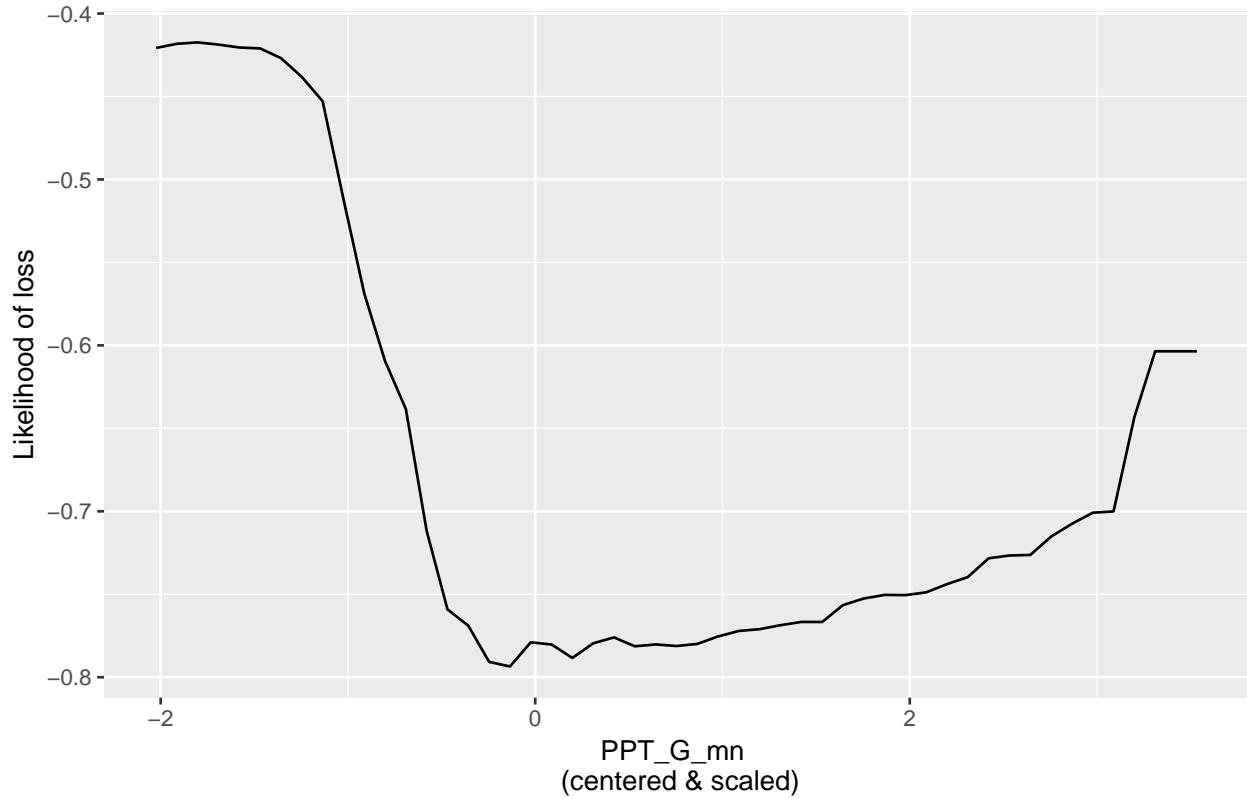
We can now look at the partial dependence plots to help understand the non-linear response of reported damages to variables of interest, such as **A\_max**.

Figure 11: Partial dependency plot for influence of **A\_max** on wheat damage prediction



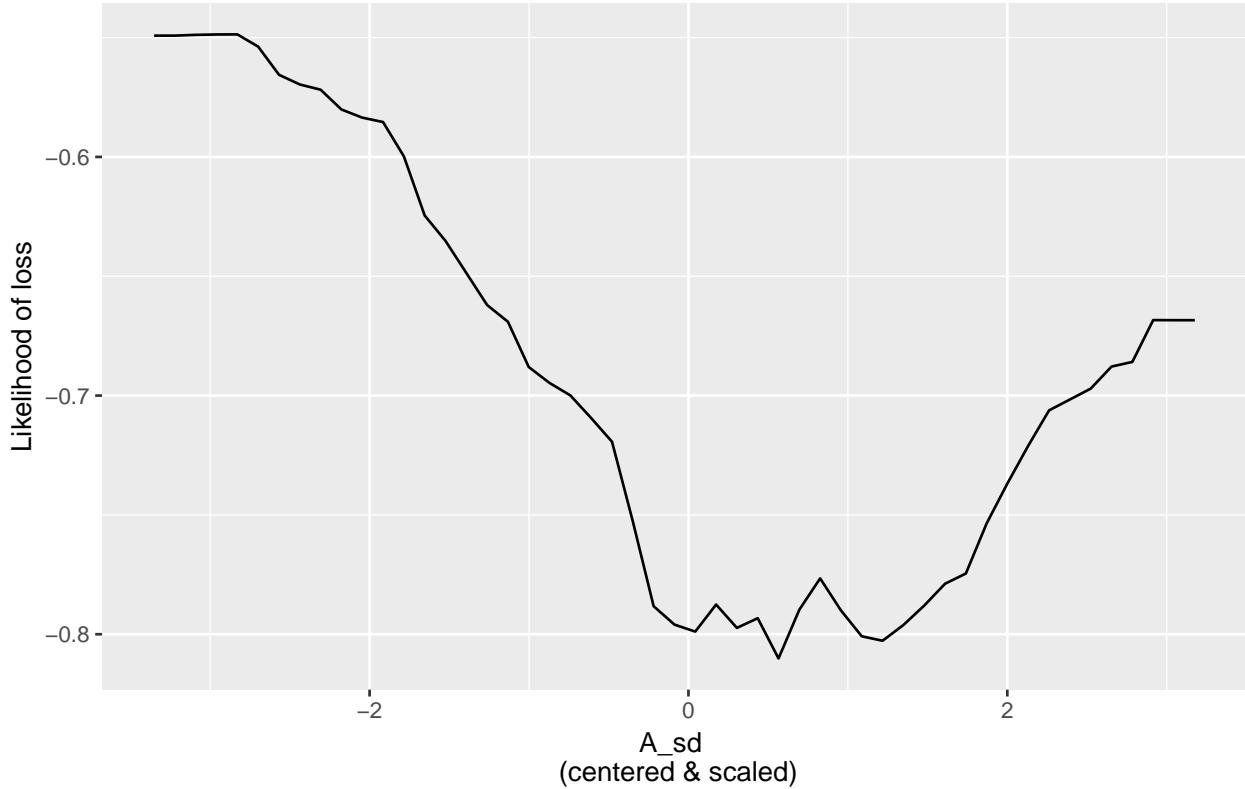
In the figure above, we can see, similar to maize, how the likelihood of reporting maize losses decreases rapidly after one standard deviation below the mean value of  $A_{\max}$ , with the likelihood of losses increasing at the upper end of the distribution. These losses at upper end of the distribution might reflect excess rainfall, or late season losses from lodging or frost. Ideally additional variables would capture the effects of these late season dynamics in ways that maximum NDVI cannot. Similarly, we can look at the partial dependence plot for other variables of interest such as PPT\_G\_mn.

Figure 12: Partial dependency plot for influence of PPT\_G\_mn on wheat damage prediction



In the figure above, we can see some interesting similarities and differences between the wheat and maize response to mean precipitation over the growing season. Of particular interest is between the mean and two standard deviations above the mean. The likelihood that substantial losses are reported doesn't increase meaningfully until over +2sd above the mean. In comparison, maize losses increased rapidly at +1sd. Past 3 standard deviation the likelihood of losses increases rapidly. Would likely be linked to losses from rust or mold, or physical damage from storms all of which increase with excess rainfall. Similarly, we can look at the partial dependence plot for other variables of interest such as A\_sd.

Figure 13: Partial dependency plot for influence of A\_sd on wheat damage prediction



In the figure above, we can see some interesting similarities and differences between the wheat and maize response to variability in greenness. Of particular interest is the broader range of variability that has no substantial influence on the likelihood of losses. This is not necessarily pointing out that wheat is more resiliant to changes, but might be an indication of a broader range of natural variability in greenness that is expected for wheat.

### Sorghum Damages

The confusion matrix compares predicted values against the observed (reference). We also report the a) overall accuracy, 2) Kappa coefficient, and 3) the recall rate<sup>6</sup> as seen in table 10 below:

*Table 10: Confusion matrix for reported damages > 30% for Sorghum*

Table 5: FALSE = Crop damages greater than 30%

	FALSE	TRUE
FALSE	1200	36
TRUE	205	3569

*Table 11: Performance metrics for reported damages > 30% for Sorghum*

	Accuracy	Kappa	Recall
Value	0.95	0.88	0.85

<sup>6</sup>the percentage of sub-kebeles with ‘substantial damage’ that are correctly classified by our models.

For the case of Sorghum, we see impressive identification of reported Sorghum crop losses with a recall rate of 0.85, indicating that 85% of all substantial loss cases were predicted in our completely independent *testing* data. In all, 1200 out of 1405 reported cases were predicted, and only 36 cases were falsely predicted as crop failure areas.

### Barley Damages

The confusion matrix compares predicted values against the observed (reference). We also report the a) overall accuracy, 2) Kappa coefficient, and 3) the recall rate<sup>7</sup> as seen in table 12 below:

*Table 12: Confusion matrix for reported damages > 30% for barley*

Table 7: FALSE = Crop damages greater than 30%

	FALSE	TRUE
FALSE	559	78
TRUE	556	3739

*Table 13: Performance metrics for reported damages > 30% for barley*

	Accuracy	Kappa	Recall
Value	0.87	0.57	0.5

For the case of barley, we see relatively poor identification of reported losses with a recall rate of only 0.5, indicating that only 50% of all substantial loss cases were predicted in our independent *testing* data. In all, 559 out of 1115 reported cases were predicted, and 78 cases were falsely predicted as crop failure areas.

### Teff Damages

The confusion matrix compares predicted values against the observed (reference). We also report the a) overall accuracy, 2) Kappa coefficient, and 3) the recall rate<sup>8</sup> as seen in table 14 below:

*Table 14: Confusion matrix for reported damages > 30% for barley*

Table 9: FALSE = Crop damages greater than 30%

	FALSE	TRUE
FALSE	816	24
TRUE	655	4463

*Table 15: Performance metrics for reported damages > 30% for barley*

	Accuracy	Kappa	Recall
Value	0.89	0.64	0.55

For the case of teff, we see relatively poor identification of reported losses with a recall rate of only 0.55,

<sup>7</sup>the percentage of sub-kebeles with ‘substantial damage’ that are correctly classified by our models.

<sup>8</sup>the percentage of sub-kebeles with ‘substantial damage’ that are correctly classified by our models.

indicating that only 55% of all substantial loss cases were predicted in our independent *testing* data. In all, 816 out of 1471 reported cases were predicted, and 24 cases were falsely predicted as crop failure areas.

“

## References

- Asrar, GQ, M Fuchs, ET Kanemasu, and JL Hatfield. 1984. "Estimating Absorbed Photosynthetic Radiation and Leaf Area Index from Spectral Reflectance in Wheat." *Agronomy Journal* 76 (2). American Society of Agronomy: 300–306.
- Bachewe, Fantu, Feiruz Yimer, and Bart Minten. n.d. *Agricultural Prices During Drought in Ethiopia: An Updated Assessment Using National Producer Data (January 2014 to June 2016)*. *Agricultural Prices During Drought in Ethiopia: An Updated Assessment Using National Producer Data (January 2014 to June 2016)*.
- Bartholome, Etienne. 1988. "Radiometric Measurements and Crop Yield Forecasting Some Observations over Millet and Sorghum Experimental Plots in Mali." *International Journal of Remote Sensing* 9 (10-11). Taylor & Francis: 1539–52.
- Becker-Reshef, Inbal, Chris Justice, Mark Sullivan, Eric Vermote, Compton Tucker, Assaf Anyamba, Jen Small, et al. 2010. "Monitoring Global Croplands with Coarse Resolution Earth Observations: The Global Agriculture Monitoring (Glam) Project." *Remote Sensing* 2 (6). Molecular Diversity Preservation International: 1589–1609.
- Becker-Reshef, Inbal, Eric Vermote, Mark Lindeman, and Christopher Justice. 2010. "A Generalized Regression-Based Model for Forecasting Winter Wheat Yields in Kansas and Ukraine Using Modis Data." *Remote Sensing of Environment* 114 (6). Elsevier: 1312–23.
- Benedetti, Roberto, and Paolo Rossini. 1993. "On the Use of Ndvi Profiles as a Tool for Agricultural Statistics: The Case Study of Wheat Yield Estimate and Forecast in Emilia Romagna." *Remote Sensing of Environment* 45 (3). Elsevier: 311–26.
- Biradar, Chandrashekhar M, and Xiangming Xiao. 2011. "Quantifying the Area and Spatial Distribution of Double-and Triple-Cropping Croplands in India with Multi-Temporal Modis Imagery in 2005." *International Journal of Remote Sensing* 32 (2). Taylor & Francis: 367–86.
- Bolton, Douglas K, and Mark A Friedl. 2013. "Forecasting Crop Yield Using Remotely Sensed Vegetation Indices and Crop Phenology Metrics." *Agricultural and Forest Meteorology* 173. Elsevier: 74–84.
- Brunsdon, Chris, A Stewart Fotheringham, and Martin E Charlton. 1996. "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity." *Geographical Analysis* 28 (4). Wiley Online Library: 281–98.
- Brunsdon, Chris, Stewart Fotheringham, and Martin Charlton. 1998. "Geographically Weighted Regression." *Journal of the Royal Statistical Society: Series D (the Statistician)* 47 (3). Wiley Online Library: 431–43.
- Butler, Ethan E, and Peter Huybers. 2015. "Variations in the Sensitivity of Us Maize Yield to Extreme Temperatures by Region and Growth Phase." *Environmental Research Letters* 10 (3). IOP Publishing: 034009.
- Clevers, JGPW. 1997. "A Simplified Approach for Yield Prediction of Sugar Beet Based on Optical Remote Sensing Data." *Remote Sensing of Environment* 61 (2). Elsevier: 221–28.
- Daughtry, CST, KP Gallo, and Marvin E Bauer. 1983. "Spectral Estimates of Solar Radiation Intercepted by Corn Canopies." *Agronomy Journal* 75 (3). American Society of Agronomy: 527–31.
- De Wit, AJW de, and CA Van Diepen. 2007. "Crop Model Data Assimilation with the Ensemble Kalman Filter for Improving Regional Crop Yield Forecasts." *Agricultural and Forest Meteorology* 146 (1). Elsevier: 38–56.
- Didan, Kamel, and Alfredo Huete. 2006. "MODIS Vegetation Index Product Series Collection 5 Change Summary." *TBRS Lab, the University of Arizona*.
- Doraiswamy, Paul C, Sophie Moulin, Paul W Cook, and Alan Stern. 2003. "Crop Yield Assessment from Remote Sensing." *Photogrammetric Engineering & Remote Sensing* 69 (6). American Society for

- Photogrammetry; Remote Sensing: 665–74.
- Doraiswamy, Paul C, Alan J Stern, and Bakhyt Akhmedov. 2007. “Crop Classification in the Us Corn Belt Using Modis Imagery.” In *Geoscience and Remote Sensing Symposium, 2007. Igarss 2007. Ieee International*, 809–12. IEEE.
- “Ethiopia: Drought - 2015-2017.” n.d. *ReliefWeb*. <https://reliefweb.int/disaster/dr-2015-000109-eth>.
- Friedl, Mark A, Douglas K McIver, John CF Hodges, XY Zhang, D Muchoney, Alan H Strahler, Curtis E Woodcock, et al. 2002. “Global Land Cover Mapping from Modis: Algorithms and Early Results.” *Remote Sensing of Environment* 83 (1). Elsevier: 287–302.
- Friedl, Mark A, Damien Sulla-Menashe, Bin Tan, Annemarie Schneider, Navin Ramankutty, Adam Sibley, and Xiaoman Huang. 2010. “MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets.” *Remote Sensing of Environment* 114 (1). Elsevier: 168–82.
- Funk, Chris, and Michael E Budde. 2009. “Phenologically-Tuned Modis Ndvi-Based Production Anomaly Estimates for Zimbabwe.” *Remote Sensing of Environment* 113 (1). Elsevier: 115–25.
- Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, et al. 2015. “The Climate Hazards Infrared Precipitation with Stations—a New Environmental Record for Monitoring Extremes.” *Scientific Data* 2. Nature Publishing Group: 150066.
- Gao, Feng, Martha C Anderson, Xiaoyang Zhang, Zhengwei Yang, Joseph G Alfieri, William P Kustas, Rick Mueller, David M Johnson, and John H Prueger. 2017. “Toward Mapping Crop Progress at Field Scales Through Fusion of Landsat and Modis Imagery.” *Remote Sensing of Environment* 188. Elsevier: 9–25.
- Gollini, Isabella, Binbin Lu, Martin Charlton, Christopher Brunsdon, and Paul Harris. 2013. “GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models.” *arXiv Preprint arXiv:1306.0413*.
- Gray, Josh, Mark Friedl, Steve Frolking, Navin Ramankutty, Andrew Nelson, and Murali Krishna Gumma. 2014. “Mapping Asian Cropping Intensity with Modis.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (8). IEEE: 3373–9.
- Groten, SME. 1993. “NDVI—crop Monitoring and Early Yield Assessment of Burkina Faso.” *TitleREMOTE SENSING* 14 (8). Taylor & Francis: 1495–1515.
- Hatfield, JL. 1983. “Remote Sensing Estimators of Potential and Actual Crop Yield.” *Remote Sensing of Environment* 13 (4). Elsevier: 301–11.
- Idso, Sherwood B, Ray D Jackson, and Robert J Reginato. 1977. “Remote-Sensing of Crop Yields.” *Science* 196 (4285). American Association for the Advancement of Science: 19–25.
- Li, Le, Mark A Friedl, Qinchuan Xin, Josh Gray, Yaohong Pan, and Steve Frolking. 2014. “Mapping Crop Cycles in China Using Modis-Evi Time Series.” *Remote Sensing* 6 (3). Multidisciplinary Digital Publishing Institute: 2473–93.
- Lobell, David B, and Gregory P Asner. 2004. “Cropland Distributions from Temporal Unmixing of Modis Data.” *Remote Sensing of Environment* 93 (3). Elsevier: 412–22.
- Lobell, David B, David Thau, Christopher Seifert, Eric Engle, and Bertis Little. 2015. “A Scalable Satellite-Based Crop Yield Mapper.” *Remote Sensing of Environment* 164. Elsevier: 324–33.
- Lobell, David B., Wolfram Schlenker, and Justin Costa-Roberts. 2011. “Climate Trends and Global Crop Production Since 1980.” *Science* 333 (6042). American Association for the Advancement of Science: 616–20. doi:10.1126/science.1204531.
- MacDonald, Robert B, and Forrest G Hall. 1980. “Global Crop Forecasting.” *Science* 208 (4445). Washington: 670–79.
- Major, JACK. 1967. “Potential Evapotranspiration and Plant Distribution in Western States with Emphasis

on California.” *AMER ASSOC Adv SCI PUBL* 86.

Mann, Michael L, and James M Warner. 2017. “Ethiopian wheat yield and yield gap estimation: A spatially explicit small area integrated data approach.” *Field Crops Research* 201. Elsevier: 60–74.

Mann, Michael L, Enric Batllori, Max A Moritz, Eric K Waller, Peter Berck, Alan L Flint, Lorraine E Flint, and Emmalee Dolfi. 2016. “Incorporating Anthropogenic Influences into Fire Probability Models: Effects of Human Activity and Climate Change on Fire Activity in California.” *PLoS One* 11 (4). Public Library of Science: e0153589.

Mann, ML L, and J Warner. 2015. “Ethiopian Wheat Yield and Yield Gap Estimation: A Small Area Integrated Data Approach.” Addis Ababa, Ethiopia: International Food Policy Research Institute.

“Mid-Year Review, Ethiopia Humanitarian Requirements Document, July 2017.” 2017. *ReliefWeb*. <https://reliefweb.int/report/ethiopia/mid-year-review-ethiopia-humanitarian-requirements-document-july-2017>.

Mkhabela, MS, Paul Bullock, S Raj, S Wang, and Y Yang. 2011. “Crop Yield Forecasting on the Canadian Prairies Using Modis Ndvi Data.” *Agricultural and Forest Meteorology* 151 (3). Elsevier: 385–93.

Mo, X, S Liu, Z Lin, Y Xu, Y Xiang, and TR McVicar. 2005. “Prediction of Crop Yield, Water Consumption and Water Use Efficiency with a Svat-Crop Growth Model Using Remotely Sensed Data on the North China Plain.” *Ecological Modelling* 183 (2). Elsevier: 301–22.

Moriondo, M, F Maselli, and M Bindi. 2007. “A Simple Model of Regional Wheat Yield Based on Ndvi Data.” *European Journal of Agronomy* 26 (3). Elsevier: 266–74.

NASA. 1984. *AgRISTARS: Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing. Technical Report Research Report. AgRISTARS: Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing. Technical Report Research Report*. NASA.

NASS, USDA. 2003. “USDA-National Agricultural Statistics Service, Cropland Data Layer.” *United States Department of Agriculture, National Agricultural Statistics Service, Marketing and Information Services Office, Washington, DC [Available at Http//Nassgeodata. Gmu. Edu/Crop-Scape, Last Accessed September 2012.]*.

Pinter Jr, Paul J, Jerry C Ritchie, Jerry L Hatfield, and Galen F Hart. 2003. “The Agricultural Research Service’s Remote Sensing Program.” *Photogrammetric Engineering & Remote Sensing* 69 (6). American Society for Photogrammetry; Remote Sensing: 615–18.

Portmann, Felix T, Stefan Siebert, and Petra Döll. 2010. “MIRCA2000—Global Monthly Irrigated and Rainfed Crop Areas Around the Year 2000: A New High-Resolution Data Set for Agricultural and Hydrological Modeling.” *Global Biogeochemical Cycles* 24 (1). Wiley Online Library.

Ramankutty, Navin, Amato T Evan, Chad Monfreda, and Jonathan A Foley. 2008. “Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000.” *Global Biogeochemical Cycles* 22 (1). Wiley Online Library.

Rasmussen, Michael S. 1992. “Assessment of Millet Yields and Production in Northern Burkina Faso Using Integrated Ndvi from the Avhrr.” *International Journal of Remote Sensing* 13 (18). Taylor & Francis: 3431–42.

Ray, Deepak K, James S Gerber, Graham K MacDonald, and Paul C West. 2015. “Climate variation explains a third of global crop yield variability.” *Nature Communications* 6 (January). The Author(s): 5989. <http://dx.doi.org/10.1038/ncomms6989> <http://10.1038/ncomms6989> <http://www.nature.com/articles/ncomms6989{\#}supplementary-information>.

Schemm, Paul. 2016. “Ethiopia Is Facing a Devastating Drought, and Food Aid May Soon Run Out.” *The Washington Post*. WP Company. [http://www.washingtonpost.com/sf/world/2016/02/22/history-repeats-itself-in-ethiopia/?utm\\_term=.2b7bb38f3a3e](http://www.washingtonpost.com/sf/world/2016/02/22/history-repeats-itself-in-ethiopia/?utm_term=.2b7bb38f3a3e).

———. 2017. “Ethiopia Is Facing a Killer Drought. but It’s Going Almost Unnoticed.” *The Washington Post*. WP Company. <https://www.washingtonpost.com/news/worldviews/wp/2017/05/01/>

[ethiopia-is-facing-a-killer-drought-but-its-going-almost-unnoticed/?utm\\_term=.ad44d46b0631](#).

Thenkabail, Prasad S, Chandrashekhar M Biradar, Praveen Noojipady, Venkateswarlu Dheeravath, Yuanjie Li, Manohar Velpuri, Muralikrishna Gumma, et al. 2009. "Global Irrigated Area Map (Giam), Derived from Remote Sensing, for the End of the Last Millennium." *International Journal of Remote Sensing* 30 (14). Taylor & Francis: 3679–3733.

Thenkabail, Prasad, Parthasaradhi GangadharaRao, Trent Biggs, M Krishna, and Hugh Turrel. 2007. "Spectral Matching Techniques to Determine Historical Land-Use/Land-Cover (Lulc) and Irrigated Areas Using Time-Series 0.1-Degree Avhrr Pathfinder Datasets." *Photogrammetric Engineering & Remote Sensing* 73 (10): 1029–40.

Tucker, CJ, and PJ Sellers. 1986. "Satellite Remote Sensing of Primary Production." *International Journal of Remote Sensing* 7 (11). Taylor & Francis: 1395–1416.

Tucker, CJ, BN Holben, JH Elgin Jr, JE McMurtrey III, and others. 1980. "Relationship of Spectral Data to Grain Yield Variation." *Photogrammetric Engineering and Remote Sensing* 46 (5): 657–66.

Wardlow, Brian D, and Stephen L Egbert. 2008. "Large-Area Crop Mapping Using Time-Series Modis 250 M Ndvi Data: An Assessment for the Us Central Great Plains." *Remote Sensing of Environment* 112 (3). Elsevier: 1096–1116.

Xiao, Xiangming, Stephen Boles, Steve Frolking, Changsheng Li, Jagadeesh Y Babu, William Salas, and Berrien Moore. 2006. "Mapping Paddy Rice Agriculture in South and Southeast Asia Using Multi-Temporal Modis Images." *Remote Sensing of Environment* 100 (1). Elsevier: 95–113.

n.d. *Data Portals*. FEWS NET. <https://earlywarning.usgs.gov/fews/downloads/index.php>.