Research paper

Ethiopian wheat yield and yield gap estimation: A spatially explicit small area integrated data approach

Michael L. Mann [a,*], James M. Warner [b]

[a] Department of Geography, The George Washington University, Washington DC, United States
[b] International Food Policy Research Institute, Addis Ababa, Ethiopia

ABSTRACT

Despite the routine collection of annual agricultural surveys and significant advances in GIS and remote sensing products, little econometric research has integrated these data sources in estimating developing nations' agricultural yields. In this paper, we explore the determinants of wheat output per hectare in Ethiopia during the 2011–2013 principal *Meher* crop seasons at the kebele administrative area. Using a panel data approach, combining national agricultural field surveys with relevant GIS and remote sensing products, the model explains nearly 40% of the total variation in wheat output per hectare across the country. Reflecting on the high interannual variability in output per hectare, we explore whether these changes can be explained by weather, shocks to, and management of rain-fed agricultural systems. The model identifies specific contributors to wheat yields that include farm management techniques (e.g. area planted, improved seed, fertilizer, and irrigation), weather (e.g. rainfall), water availability (e.g. vegetation and moisture deficit indexes) and policy intervention. Our findings suggest that woredas produce between 9.8 and 86.5% of their locally attainable wheat yields given their altitude, weather conditions, terrain, and plant health. In conclusion, we believe the combination of field surveys with spatial data can be used to identify management priorities for improving production at a variety of administrative levels.

1. Introduction

1.1. The Ethiopian context

Ethiopia's agriculture system constitutes 46% of gross national production, employs 85% of its population, and creates 75% of export commodity value (FDRE, 2013). Despite its large scale, the agricultural sector is largely formed by smallholder subsistence farms burdened by dependence on erratic rain-fed systems. In all, smallholders account for 96% of total area cultivated (Taffesse et al., 2011). Ethiopia's rain dependent agricultural system is particularly vulnerable to shifts in climate and weather, with less than 3% of households having access to irrigation (or less than 1% of cereal acreage) (Mann and Warner, 2015; Taffesse et al., 2011). These vulnerabilities are further exaggerated by extensive use, land degradation, and household poverty.

In Ethiopia, a variety of climate extreme events are increasingly common, particularly droughts and floods. Changes in weather and climate, especially in the context of the dynamic Sahel monsoon, form a potential threat to agricultural production and food security throughout the region. Recent evidence suggests that the incidence of droughts and floods in Ethiopia has increased in the last ten years relative to the decade before (FDRE, 2013). Drought events alone are estimated to reduce Ethiopia's GDP by up to 1% in a typical year (FDRE, 2013). This will likely be confounded by additional loss of agricultural productivity due to changes in climate (Jones and Thornton, 2003). Targeted intervention can lead to increases in yields in some of Ethiopia's most challenging environments (Mann and Warner, 2015). Despite the critical nature of this research, little is known about the response of smallholders to these trends, especially across large spatial scales and across heterogeneous physical and social terrain (Altieri and Koohafkan, 2008).

Wheat is modeled here due to its relative importance as well as its wide scale adoption throughout the four main regions of Ethiopia. According to recent estimates there are approximately 4.7 million farmers growing wheat on approximately 1.6 million hectares representing between 15 and 18% of total crop area (Minot et al., 2015). Additionally, less than 1% of all wheat production takes place outside the four regions studied in this paper.

1.2. Data and integration

One of the primary objectives of this study is to integrate data from a variety of sources to better model the effects of weather, climate, markets, and farm management behavior on agricultural productivity. Each source provides a unique and often complimentary set of information. For instance, every year Ethiopia's Central Statistics Authority (CSA) engages in the massive Agricultural Sample Survey (AgSS) of over 45,000 rural households. This survey provides critical information on crop yields, and management practices at the kebele level. Although the AgSS is over ten times larger than any other agricultural survey in the nation, even moderate resolution remotely sensed imagery observes around 44 million locations in Ethiopia daily. The fusion of these two datasets can therefore leverage the household-level detail provided by AgSS with nearly real-time information provided by remote sensing. Combined, they may reveal new information relevant to researchers, policy makers, and optimally the farmers themselves.

Although there are individual benefits and drawbacks to both, combining survey and remotely sensed data can enhance yield analysis. For example, conventional agricultural surveys have subjectivity in responses, and implementation can be expensive and time-consuming; meanwhile remote sensing can provide objective, standardized, and possibly cheaper information in a more timely fashion to aid farmers through monitoring, as well as yield predictions (Ahmad et al., 2014; Lobell et al., 2005). Surveys provide detailed information on farm labor, input choice, extension access, and other key determinants of productivity. On the other hand, remote sensing is subject to scale issues (especially in areas of small farmer plots with extensive multi-cropping), cloud cover during the rainy season, methodological choices by the researcher (eg. estimated planting time) as well as other drawbacks. Both methods have weaknesses and strengths, but by combining both, we can exploit a broader and more flexible set of data to model the determinants of wheat productivity. For instance, low yields observed from crop cut surveys might be explained by changes in rainfall patterns easily observed by satellite, alternatively while satellite imagery provides a real-time look at plant health, it can't determine whether shifts are the result of pests or disease.

1.3. Project description

Despite routine agricultural surveys and advances in agricultural modeling, little effort has been made to understand the determinants of agricultural yields (and gaps) through spatial data integration methods (Lobell et al., 2005). Here we aim to see if our understanding of the determinants of Ethiopia's agricultural productivity can be improved through the integration of data between disciplines and across space. This study combines four basic types of data: remote sensing data, policy intervention information, agricultural household surveys, and other spatial data such as information on road networks and edaphic properties. We combined this information to provide two products that are nationally representative 1) wheat productivity at the kebele level,[1] and 2) yield gaps at the woreda level for wheat in Ethiopia's four major growing regions. While this form of analysis has been developed in more economically developed nations with larger agricultural crop planting areas and more homogenous agroecological climates (Fontana et al., 2005; Randall et al., 2011), there has been relatively little work focusing on the African smallholder farmer.

Current academic and applied work in remote sensing is generally assessed at the pixel level, either at the macro (ie. global, regional, state) or local scale (individual field) (Ferenz et al., 2004; Liu et al., 2005; Prasad et al., 2006). Global studies of agriculture rely on broad remote sensing tools, and the findings are, understandably, across multiple countries and agroecological zones (e.g. Licker et al., 2011). Field-level studies rely on localized agricultural plots (typically of large multi-hectare fields) and are combined with spatially consistent satellite imagery that create relatively accurate yield estimates (Ferenz et al., 2004; Serrano et al., 2000). Usually, these farm-level research projects have taken place in more developed nations where agricultural plots are significantly larger and monocropped (Moran et al., 1997; Swinton and Lowenberg-Deboer, 2001). To our knowledge, the type of national level analysis presented here, particularly for African smallholder farmers, has not been performed to date. In addition, most research of this kind has relied on crop models that view modeling from a purely agroecological perspective.

We believe models estimating observed yields may be improved by expanding their scope to include variables that better capture the idiosyncratic nature of actual management practices, spatial patterns of investment and market activity, and other critical spatial determinants. Here we develop crop-specific national-level productivity maps for wheat that are beneficial for understanding overall productivity as well as evaluating research and policy interventions. The use of a broad set of spatial variables, integrated with household surveys, allows for increased flexibility and a better understanding of what influences Ethiopian farmers' yields—estimates recorded by households in three annual nationally representative surveys. This study provides time variant estimations of crop productivity for Ethiopia using both survey and spatial data. We also present estimates of wheat yield 'gaps', where 'gaps' are defined as the difference between average yields and 90th percentile locally-attained yields from areas of similar climate, soil, and water availability. These estimates are critical to understand the production potential of the agricultural sector as well as to guide interventions, both locally and nationally. In this way, we hope to provide both researchers and policymakers with improved information to enhance analysis and interventions across Ethiopia's diverse agrological landscape.

2. Methods

2.1. Objectives and overview

We develop a panel regression model to estimate 1) wheat output per hectare, to produce 2) yield gaps for three seasons including the 2012–2014 Meher crop seasons in Ethiopia. This study focuses on the four major agricultural regions of Ethiopia (Tigray, Amhara, Oromia, and SNNP) because they comprise the majority of population and agricultural crop production for the country. The four regions have approximately 50 zones, 550 agricultural woredas, and 14,500 kebeles. Even though about 2150 of these kebeles are surveyed per year in AgSS, for stability of estimates, we restricted our sample to those kebeles where three or more wheat farmers were identified. Additionally, while this paper estimates at the sub-woreda, or kebele level, most operational policy is implemented within woredas and for that reason our estimates are aggregated to this level.

2.2. Model data

The full model uses 21 independent variables that we divide into Survey, Climate/Weather, Spatial, and Remotely Sensed data. In order to match with AgSS survey variables, all explanatory vari-

---

[1] The levels of administrative areas in Ethiopia consist of, in descending order, regions, zones, woredas and kebeles. There are an average of 15 zones per region in the four major wheat producing regions, with approximately 11 woredas per zone. Rural woredas are composed of approximately 24 wards/kebeles each, and are the usual focus when speaking of local decentralized government administration.

ables are summarized by the mean or standard deviation[2] at the *kebele*-level.

2.2.1. Survey variables

Primary data are obtained from Ethiopia's annual Agricultural Sample Survey (AgSS) and the Population Census Commission. Survey data includes productivity and farm management variables covering three years of agricultural data (2011/2012–2013/2014 Meher crop seasons[3]) (AgSS, 2014, 2013, 2012). The CSA takes a random sample of enumeration areas (EAs[4]) that exist at the sub-*kebele* level. *Kebele* population estimates from the 2007 Ethiopian census are used to weight *kebele*-level agricultural statistics (Population Census Commission, 2008), and adjusted over time using the World Bank's annual agricultural population growth rates for Ethiopia (The World Bank, 2013). We treat the 20 households interviewed per EA as being a representative sample of the *kebele*. Importantly, approximately 97% of the same EAs are sampled for a three-year period. This allows for a *kebele* panel data set to be used. Field productivity is measured as quintals per hectare and is a sample of between three and five crop cuts at the EA level.[5] Observations with less than three crop cuts per EA were dropped from the analysis to enhance stability of the estimates. Population density is measured as the total population per hectare (*Pop_Den*). Other statistics from the AgSS on wheat production at the *kebele level* include the proportion of land applied with chemical fertilizer (*Chemfert*); the proportion of land where wheat was irrigated (*Irrigation*); the proportion of land area dedicated to improved seed (*ImprSeed*); the proportion of land where wheat was damaged by weather, pests or other external events (*Damage*); and the proportion of total area planted with wheat (*LandWheat*). To control for the effects of the adoption of knowledge-based technologies our regression also includes the temporally lagged value of WheatOPH.

2.2.2. Climate/Weather variables

In order to capture the influence of longer-term weather expectations, two measures of climatic water deficit (CWD) were obtained from hydrologic models (Willmott and Matsuura, 2001; Willmott, 1977). CWD is an approximation of water available to plants, as it expresses the annual evaporative demand that exceeds available water. CWD integrates complex interactions of heat, water supply and demand, terrain, and edaphic properties (e.g. soil water holding capacity) into simple measures of seasonal characteristics relevant to plant community structure and productivity through a measure of water stress. *CWD* data was retrieved at 0.5° resolution (a pixel equals 0.5 decimal degrees or approximately 50 km on a side) (Willmott and Matsuura, 2001). A simplified visual representation is presented in Appendix A (Fig. A1). The variables used here include a measure of the mean (*CWDMN*) and standard deviation (*CWDSD*) of CWD for the 1981–2010 period. These act as a proxy for long-term expectations of uncertainty around water

availability for a given location. Estimates of CWD were not available after 2010.

2.2.3. Spatial variables

Nine variables were obtained to characterize the geographic locations, terrain, soils and other determinants of productivity. Administrative boundaries were downloaded from UN-OCHA's *Humanitarian Response COD-FOD Registry* (UN-OCHA, 2014). Euclidean distance to Addis Ababa is calculated in meters (*Dist_Addis*), road density is estimated as road length as described by the Ethiopian Water and Land Resource Centre (WLRC, 2013) in length of road within 5 km of any given pixel, in km per km² (*Road_Den*). Agroecological zones (*AgroEco*, dummy variables) were obtained from the Water and Land Resource Centre (WLRC, 2013) and describe areas of similar growing conditions, with 15 zones based roughly on elevation and average precipitation. *AgroEco* therefore controls for omitted variables correlated with differences between these zones. Elevation in meters (*Elev*) was obtained from the Shuttle Radar Topography Mission (STRM) at 90 m resolution (Jarvis et al., 2008) and was used to calculate terrain slope in degrees (*Slope*). Cation-exchange capacity (*CEC*) was measured as the maximum quantity of total cations measured in meq+/100 g dry soil, that a soil is capable of holding, at a given pH value, available for exchange with the soil solution. CEC was used as a measure of fertility and nutrient retention capacity. This data was accessed via the Africa Soils Profiles Database v 1.1 (Leenaars, 2013). Other edaphic properties were also downloaded from the Africa Soils Database but were dropped from the final model due to statistical insignificance. The Agricultural Growth Program (*AGP*) interventions variable indicate whether or not the AGSS surveyed-areas operated under the AGP program during the sample period. The AGP intervention is a large-scale (83 initial *woredas*) project, funded by the World Bank, designed to increase productivity and marketization within high-potential agricultural areas. Intervention *woredas* were obtained from the Ethiopia Strategy and Support Program II's baseline report (IFPRI, 2013). To capture larger-scale administrative impacts, a dummy variable was used for each zone (*Z_Code*). Finally, to test for potential spatial autocorrelation locational coordinates, the centroid, were used for each *woreda* (*Cen_Y* and *Cen_X*).

2.2.4. Remotely sensed data

We use three sources of remotely sensed information from the National Aeronautics and Space Administration (NASA). Data on precipitation (*PRECIP*), was collected from the adjusted merged-IR precipitation product from Tropical Rainfall Measuring Mission (TRMM), which is collected every three hours at 0.25° resolution (25 km on a side) for the whole period of interest, and is measured in mm/h (GSFC, 2014). Annual totals were calculated for each sample year in cumulative mm for the period of the AgSS survey. As such, the TRMM data provides a real-time look at precipitation.

Remotely sensed vegetation indexes were obtained from the 16 day MODIS MOD13Q1 composite product[6] at 250 m resolution (Didan and Huete, 2006). MODIS's Enhanced Vegetation Index (EVI) is sensitive to the amount of chlorophyll in any given pixel. EVI and similar indexes are commonly used to estimate plant productivity and health in agricultural applications. After preprocessing and the identification of agricultural pixels, a series of EVI statistics were produced for each year of the AgSS survey. Refer to the *Enhanced Vegetation Index* in Appendix A for detailed methodology concerning preprocessing and the identification of agricultural pixels. EVI statistics used in the final model include *EVIMX*, the maximum achieved annual EVI value, and *EVID* the annual total area

---

[2] Standard deviations are used on a limited set of variables in order to capture interannual variability.

[3] The *Meher* crop season produces over 95% of total crop production annually (AgSS, 2013).

[4] Enumeration areas are sample areas, designed by CSA, that exist within kebele boundaries and consist of approximately 150–200 households. On average, there are about 4.8 EAs per kebele but this varies significantly because of differences in population.

[5] CSA's methodology concerning crop productivity determination does not allow for individual estimates and all farmers are projected to have the same output per hectare at the sample level via an average crop cut. More specifically, CSA takes a crop cut sample of up to a total of five farmers, depending on how many farmers actually grow the crop, and projects that average to all land dedicated to that particular crop in the sample. Therefore, this sample design allows for only one productivity estimate per EA.

[6] Composite images compile data over a fixed number of days (typically 8 or 16 days) to create cloud-free images representative of that period.
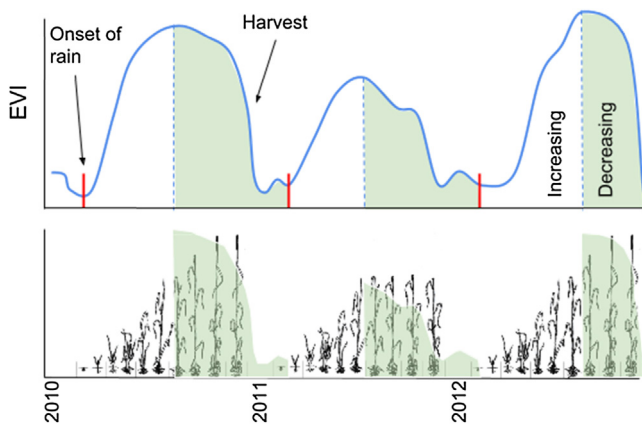
Fig. 1. Example relating EVI time series to wheat yields.
A stylized comparison of the evolution of the enhanced vegetation index (top) and plant development (bottom) under good and poor plant health. The top panel also points out key portions of the phenological cycle including the onset of the Meher planting season (red dash), maximum EVI values (blue vertical dashed line), and the decreasing portion of the EVI curve (labeled 'decreasing').
Adapted from (FAO, 2013).

under the curve of the decreasing portion of the EVI curve (green shaded areas labeled "decreasing" in Fig. 1).

To identify key points in phenology, we assume that EVIMX can be represented as the maximum EVI value for a given year and that this occurred during the *Meher* growing season. Because Ethiopia has two growing seasons, the area under the curve estimates required that measurements start at the beginning of the *Meher* growing season. We define the start of the *Meher* growing season as the annual local minimum nearest to the *Meher* rainy season onset (red vertical dash, Fig. 1), where the day of rainy season onset can be any date between May and the end of August depending on geographical location. On the basis of local expert knowledge, the start of the rainy season is estimated to occur in the first ten-day period that accounts for 2.5% of total annual rainfall, that is followed by a 30-day period that experiences no more than seven consecutive days without rain as estimated from the TRMM precipitation data[7] In addition, we focus on the decreasing portion of the EVI signal past the annual EVI maximum point. This period of the phenological cycle matches roughly with head development and yield formation in the plant. Therefore, inadequate water availability during these periods have been shown to strongly affect yields (FAO, 2013). Descriptive statistics for each independent variable can be found in Table 1.

2.3. Statistical methods

2.3.1. Wheat yield model estimation

We explore a variety of determinants for wheat output per hectare in the 2011–2013 *Meher* crop seasons by using a panel data approach at the *kebele* level. The use of panel data in this study alleviates two problems, unobserved spatial and temporal dynamics and homogeneity (lack of variance). If pooled together, the integration of an unbalanced statewide set of *kebeles* (N = 1726) over the 2011–2013 period (T = 3) amounts to a high degree of observed variance over both space and time.

Non-linear specifications are controlled for using cubic splines. A spline function is a smoothly joined piecewise polynomial of any degree. Smoothing is enforced by constraining functions to be continuous at knots, or join points (Durrleman and Simon, 1989).

---

[7] This definition of rainy season onset was obtained from a CIMMYT wheat expert.

Increasingly non-linear forms are only applied if a joint *t*-test, that all coefficients are equal to zero, is rejected. When referenced in the text all variables using splines will report the significance of this joint *t*-test. We indicate the number of coefficients estimated for each variable x with $x^{(Sn)}$ where S indicates the use of a spline and n indicates the number of coefficients estimated.

We estimated a fixed effect model that examines individual differences in intercepts, assuming the same slopes and constant variance across individual observations (group and entity) in Eq. (1):

$$\ln(y_{it}) = (\alpha + \mu_w) + \beta A_{it} + \delta W_{it} + \omega M_{it} + \varphi P_{it}$$
$$+ \gamma I_{it} + \rho P_{it} + \theta \ln(y_{it-1}) + v_{it} \tag{1}$$

Where y_{it} is the natural log of wheat output per hectare (WheatOPH) for *kebele* i in time t. βA_{it} is a Kx1 vector of regression coefficients (β) for descriptive variables A_{it}, where K is the number of descriptive variables. βA_{it} includes relevant exogenous agroecological determinants of wheat productivity ($EVIMX^{(S1-3)}$, $EVID^{(S1-3)}$, $Elev^{(S1-3)}$, $Slope^{(S1-2)}$, CEC, AgroEco), δW_{it} includes weather and climate ($CWDMN$, $CWDSD$, $Precip$, $Damage$), ωM_{it} includes management variables ($LandWheat^{(S1-2)}$, $ImprSeed^{(S1-3)}$, $ChemFert$, $Irrigation$), φP_{it} contains a policy and administration variable (AGP), γI_{it} represents infrastructure variables ($Dist\_Addis^{(S1-2)}$, $Road\_Den^{(S1-2)}$), and ρP_{it} controls for effects of population ($Pop\_Den^{(S1-2)}$). μ_w is a vector of zone-level fixed effects constants that controls for unobserved characteristics of each zone.[8] $\theta \ln(y_{it-1})$ is the temporally lagged values of y for period $t - 1$ (i.e. last year's output per hectare in the same *kebele*), and v_{it} is a $N \times T$ matrix of disturbances. All variables are standardized to have a zero mean and a standard deviation of one. As such, coefficients can be interpreted as a "one standard deviation increase in variable x corresponding to a $\hat{\beta}$ unit increase in Y", where $\hat{\beta}$ is any estimated coefficient.

Controlling for both time and zonal fixed effects allows us to disentangle the effect of temporal and spatial heterogeneity of omitted variables. To control for potential heteroscedasticity, we estimate Eq. (1) using Huber-White sandwich estimators (Huber, 1967; White, 1980). The statistical significance of each variable is tested and reported in the text as a p-value with estimates with p < 0.10 being considered statistically significant, and p > 0.10 being insignificant. Because our regression is in log-linear form we interpret the coefficients for continuous variables as percentage change in y, calculated as: %$\Delta y = 100(e^{\hat{\beta}_i} - 1)$, where $\hat{\beta}_i$ is an estimated coefficient. We present a set of regressions each exploring the explanatory power of a set of variables. These include 1) "In-Sample 1" an initial specification testing the significance of all variables of interest, 2) "In-Sample 2" a simplified model including all significant variables, and 3) "Out-of-Sample" omits all variables related to the AgSS survey, in order to allow for out-of-sample yield estimates.

2.3.2. Model tests

Pre-estimation: In order to choose between Pooled, Fixed Effects (FE), Random Effects (RE) models, an F-test, Breusch-Pagan LM test, and the Hausman Test were used. We tested if individual specific variance components were zero ($H_o : \sigma_u^2 = 0$) with the Breusch-Pagan Lagrange multiplier (BPLM) test adjusted for unbalanced

---

[8] A recent research study of yields at the global-level acknowledged that crop yield patterns do follow administrative boundaries (Licker et al., 2011). This points to the relative importance of nationally directed crop management practices over just biophysical areas determining relative yields. The implementation of many government agricultural initiatives are coordinated at the zonal administration level and then implemented in the woreda. Therefore, we incorporate zonal level administrative dummy variables to depict potential variations in administrative capacity in facilitating agricultural implementations and other potential institutional issues.

Table 1
Variable units and summary statistics (aggregated to the *kebele* level).

	Variable	Description	Units	Mean	Std.	Source
Dependent Variable	wheatOPH	Wheat output per hectare	Quintals/Hectare	17.00	8.47	Survey
Weather & Climate	CWDMN	Mean climatic water deficit mm	mm	1.49	1.86	Remotely Sensed
	CWDSD	Standard deviation climatic water deficit	mm	1.84	1.44	Remotely Sensed
	Precip	Precipitation	mm	370.46	191.70	Remotely Sensed
	Damage	Proportion of wheat damaged by weather, pests etc.	% as decimal	0.14	0.18	Survey
Management	LandWheat	Proportion of land planted with wheat	% as decimal	0.08	0.12	Survey
	ImprSeed	Proportion of land planted with improved seed	% as decimal	0.07	0.18	Survey
	ChemFert	Proportion of wheat applied with chemical fertilizers	% as decimal	0.55	0.42	Survey
	Irrigation	Proportion of wheat irrigated	% as decimal	0.01	0.05	Survey
Agroecological	EVIMX	Max enhanced vegetation index (EVI)	–	0.35	0.09	Remotely Sensed
	EVID	Area under curve of declining portion of EVI curve	–	5.08	1.79	Remotely Sensed
	Elev	Terrain elevation	Meters	1922.84	552.43	GIS
	Slope	Terrain slope	Degrees	8.37	5.16	GIS
	CEC	Soil Exchangable cations	meq+/100 g	41.68	16.16	GIS
Populations	Dist_Addis	Euclidean Distance to Addis Ababa	Kilometers	283.11	128.10	GIS
	Road_Den	Mean length of road (km) within 5 km	Meters	0.18	0.23	GIS
	Pop_Den	Population density	Person/per sq. km.	349.73	573.46	Survey

Variables *Cen_Y, Cen_X, AGP,* AgroEco and *Z-Code* are not included.

panels (Sosa-Escudero and Bera, 2008). The adjusted BPLM test does not reject the null (p > 0.08) and indicated that the RE model is not able to handle heterogeneity better than pooled OLS (i.e. inconsistent). Fixed effects was determined to be statistically significant at the 99% level (p < 0.01). The use of FE was reinforced with the use of the Hausman test implemented as a test of overidentifying restrictions (Wooldridge, 2002), where the individual effects are uncorrelated with a regressor in the model. Here, we rejected the null (p < 0.01) and concluded that individual effects μ_i are significantly correlated with at least one regressor. Therefore, the use of random effects is problematic.

Post-estimation: We tested for global spatial autocorrelation in the error term with the use of an augmented Moran's I Test for model residuals (Cliff and Ord, 1981). We do not reject the null of *no* spatial autocorrelation in the residual (p > 0.31).

2.3.3. Out-of-sample performance

The use of environmental information along with geographic and remotely sensed data enabled us to estimate wheat output per hectare out-of-sample (outside of the original AgSS survey). In order to predict *kebeles* outside of the AgSS survey, survey-related variables were dropped from Eq. (1) with predictions being made from the simplified regression (results in Table 2 – *Out-of-Sample*). These can be used to estimate yields on a *kebele*-by-*kebele*, and year-by-year basis. We used a k-fold cross-validation on the final model to evaluate the model's ability to fit out-of-sample data (Lachenbruch and Mickey, 1968). Accuracy metrics are then reported for each out-of-sample prediction in Appendix A, Table A1.

2.3.4. Yield gap methodology

Attainable yields were estimated from actual observations in Ethiopia, rather than potential yields based upon idealized or simulated conditions. Here we assume that within comparable agroecological areas the 90th percentile *kebele* is the locally attainable yield.[9] which can be contrasted with actual yields of the other

Table 2
Regression on log of output per hectare (standardized coefficients).

	In-Sample 1	In-Sample 2	Out-of-Sample
Constant	2.643e + 00**	2.593e + 00**	2.740e + 00**
LnwheatOPH$_{t-1}$	7.748e-02**	7.992e-02**	
CWDMN	1.637e-01*	1.643e-01*	8.127e-02
CWDSD	−1.956e-01**	−1.933e-01**	−1.351e-01**
Precip	−2.537e-03		
Damage	−9.376e-02**	−9.546e-02**	
EVIMX$^{(S1)}$	7.693e-02*	7.169e-02*	8.928e-02**
EVIMX$^{(S2)}$	−1.772e-01*	−1.563e-01*	−2.607e-01**
EVIMX$^{(S3)}$	9.382e-02+	7.945e-02	1.557e-01**
EVID$^{(S1)}$	1.437e-01*	1.769e-01**	1.864e-01**
EVID$^{(S2)}$	−5.840e-01*	−6.590e-01*	−8.414e-01**
EVID$^{(S3)}$	4.426e-01+	4.974e-01*	6.403e-01**
LandWheat$^{(S1)}$	1.738e-01*	1.693e-01*	
LandWheat$^{(S2)}$	−9.231e-02	−8.717e-02	
ImprSeed$^{(S1)}$	−1.621e-03	−2.206e-03	
ImprSeed$^{(S2)}$	9.093e-02	8.742e-02	
ImprSeed$^{(S3)}$	−1.306e-01	−1.262e-01	
ChemFert	3.340e-02+	3.332e-02+	
Irrigation	4.658e-04		
Elev$^{(S1)}$	8.770e-01**	8.629e-01**	9.424e-01**
Elev$^{(S2)}$	−1.233e + 00**	−1.212e + 00**	−1.100e + 00**
Elev$^{(S3)}$	6.119e-01**	6.008e-01**	4.468e-01*
Slope$^{(S1)}$	−6.786e-02+	−6.376e-02	−2.194e-02
Slope$^{(S2)}$	9.570e-02**	8.133e-02*	6.498e-02*
CEC	−2.220e-02		
AgroEco	–	–	–
Dist_Addis$^{(S1)}$	−1.806e-01**	−1.549e-01*	−1.653e-01**
Dist_Addis$^{(S2)}$	1.857e-01*	1.704e-01*	1.744e-01**
Road_Den$^{(S1)}$	4.555e-02		
Road_Den$^{(S2)}$	−7.157e-02		
Pop_Den$^{(S1)}$	2.055e-01*	2.119e-01*	
Pop_Den$^{(S2)}$	−2.641e-01**	−2.738e-01**	
AGP	4.318e-02	4.267e-02	5.241e-02*
Cen_Y	−6.190e-03		
Cen_X	−9.940e-02		
YEAR	2.037e-03		
Z_Code	–	–	–
R^2	0.39	0.38	0.24
N	1726	1726	2907

+90%, *95%, **99% level of significance.

---

[9] Given the national coverage and large variation in agroecological characteristics, determining yield potential and water-limited yield were beyond the scope of this project. We therefore relied on the locally attainable yields and chose the 90th percentile to avoid potential problems with statistical noise and aggregation over a relatively small sample of famers at the kebele level. The estimates seem relatively consistent with other Ethiopian wheat efficiencies estimated in the literature (See Alemu et al., 2014; Gelaw and Bezabih, 2009; Yami et al., 2013).

similar areas. Areas with comparable agroecological conditions are determined with a time-variant clustering algorithm.

Eight *kebele* clusters, with similar agroecological characteristics, were identified using a multidimensional clustering algorithm. The K-means algorithm partitions multidimensional data into k clusters of similar data (k = 8). Distance to the multidimensional cluster

mean determines a *kebele's* membership in any cluster each year. More specifically, given a set of *kebele* level observations ($\mathbf{x}_1$, $\mathbf{x}_2$,.., $\mathbf{x}_n$) where each observation is a multidimensional vector (containing information about climate, weather etc.), the K-mean algorithm aims to partition the *n* observations into *k* sets $\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$ in order to minimize within the cluster variance (sum of squares) as described by:

$$\min_s \sum_{i=1}^{k} \sum_{x \in S_i} |x - \rho_i|^2 \qquad (2)$$

Where ρ_i is the mean value of observations in cluster S_i and x is a vector of agroecological data for a given *kebele*. Here we partition the data based on five weather and agroecological categories. These variables include: elevation (*Elev*), precipitation (*Precip*), greenness during the late season (*EVID*), mean climatic water deficit (*CWDMN*), and standard deviation of climatic water deficit (*CWDSD*). These variables were chosen as key determinants from the estimation of Eq. (1) and can be considered representative of climate, weather, and topography. Eq. (2) is estimated as a panel with cluster membership varying year to year given changes in time-variant conditions (precipitation, greenness).

Yield performance of any *kebele* can then be compared to the distribution of yields from *kebeles* with similar climate, weather, and terrain on a year-to-year basis. Wheat yield gaps (Y_G) are calculated as follows:

$$Y_G = -100 \left(1 - \frac{Y_A}{Y_L}\right) \qquad (3)$$

Where Y_A is an actual yield and is estimated from the panel regression in Eq. (1), and Y_L is a local attainable yield, defined as the 90th percentile of the distributions described by a *kebele's agroecological* cluster membership described above. To match the scale of domestic agricultural planning, *kebele* level estimates are aggregated to the *woreda* level using an area weighted sum.

3. Results

3.1. Wheat output per hectare

The following section outlines the results of the panel regression by estimating the log of wheat output per hectare. Regression estimates are reported in Table 2, and estimates of output per hectare are presented in Fig. 2 below. The determination of the final model is shown below in Table 2 (*In-Sample 1–2*). Results for the out-of-sample estimation (for estimation of non-AgSS sampled communities) are presented in Table 2 *Out-of-Sample*. Coefficients display the expected sign and statistical significance and an overall R-square of almost 0.40 for the model.[10] Unless otherwise stated, results presented throughout the document will be from the *In-Sample 2* specification.

As outlined in Eq. (1) we test for the independent effects of agroecological conditions, weather and climate, spatial determinants, policy and administration as well as other inputs. This model also tests for the effects of improved farm management techniques including the use of improved seeds, chemical fertilizer, reported crop damage, area dedicated to wheat, as well as irrigation. As mentioned earlier, this section will refer to the results from the specification *In-Sample 2*. The results strongly support integrating survey and remotely sensed data, with most variables statistically significant at the 95% level or better and non-linear fits demonstrating expected signs.

As expected, the survey variables confirmed many of the contributing factors to wheat productivity in the final model (In-Sample 2). We find that higher yields in the previous season are correlated with increases in the current *Meher* season ($lnwheatOPH_{t-1}$). It should be emphasized that the relative size of the coefficient is not as large as might be expected and the impact of omitting the variable does not substantially alter the R-square. Specifically, a one quintile increase in last year's OPH has an approximate 0.42 quintile increase in the current output per hectare.

We find that at the *kebele* level chemical fertilizers increase production per hectare, although at current rates of application its effects are relatively small. A one standard deviation increase (ie. 42% increase in land applied with chemical fertilizer in a *kebele*) in the percentage of land applied with chemical fertilizers (*Chem-Fert*), at the average rate of application, increases wheat output per hectare by 3.38% (joint F-test, $p < 0.05$).[11] At the *kebele* level, improved seeds (*ImprSeed*) have a statistically significant effect on productivity (joint F-test, $p > 0.05$).

We also find that irrigation (*Irrigation*) has no significant effect on output per hectare ($p > 0.98$, In-Sample 1). This finding, however, is likely due to the extremely low percentage of households with irrigation (less than 3% of farmer's have access). None of the *kebeles* have enough irrigated land to see broad increases in productivity. On the other hand, damage to a percentage of holdings (*Damage*) from pests, rust, flooding, hail et cetera significantly reduces output per hectare in approximately half of all farms growing wheat ($p < 0.01$). A one standard deviation (18%) increase in the percentage of land damaged corresponds to a 9.10% decrease in output per hectare. We also saw significant increases in yield in *kebeles* where farmers plant a larger proportion of wheat (*LandWheat*) with a one standard deviation increase (12.14%) in the percentage of land planted increases output per hectare by 18.45% ($p < 0.01$).

Two climate variables, obtained via the crop model estimation, significantly contribute to explaining relative wheat productivity. We tested for climatic influences with the mean and standard deviation of climatic water deficits (*CWDMN* and *CWDSD*, respectively). Splines from *CWDMN* indicate that yields are higher in areas with historically moderate levels of available water ($p < 0.05$). The failure to find declining productivity in the driest areas is likely due to the exclusion of the driest portions of the country (outside of the four major growing regions) from the sample. Areas with higher historical variability in water availability (*CWDSD)* see declining productivity ($p < 0.01$) as relative long-term variability of rain patterns should have a negative impact on yields.

Two of the three variables related to remote sensing, are significant indicators of wheat productivity. Precipitation was found to be statistically insignificant (*Precip*). The explanatory power of precipitation (*Precip*) is undermined ($p > 0.95$, In-Sample 1) by the inclusion of variables related to *CWD* and *EVI* that are probably better proxies for water availability for plants. We control for plant health through critical periods of the growing season through the use of the Enhanced Vegetation Index (EVI). EVI maximum values (*EVIMX*) and area under the declining portion of the EVI curve of (*EVID*) both significantly correlate with output per hectare ($p < 0.07$ and $p < 0.05$, respectively). In the section below, the non-linear effects of these environmental variables are briefly outlined.

The effects of the maximum value of EVI (*EVIMX)* remains positive until very high levels and then decreases towards zero influence at the highest recorded levels (Fig. 3a). As such, *EVIMX* is capturing increases in productivity due to total growth and leaf area until this declines with the wettest areas of the south, which have extremely high EVI values and lower wheat productivity that

---

[10] Given the relative stability of area planted, we can capture nearly 75% of variation in total wheat output (quintals) at the *kebele* level using similar models. For the purpose of brevity, these models are not presented here.

[11] For all non-linear specifications, we reported the joint *t*-test that all coefficients are all equal to zero.
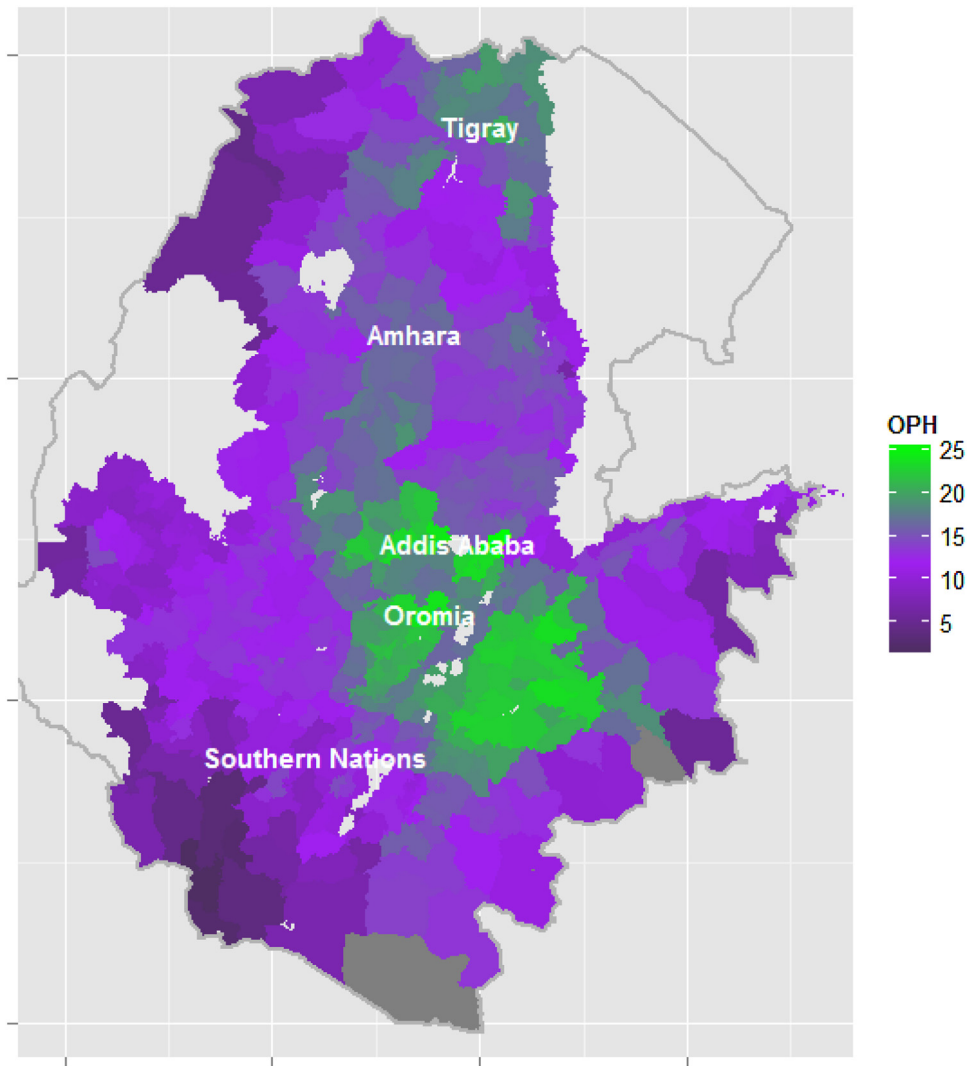
Fig. 2. Estimated wheat output per hectare in quintals by woreda.
Estimated mean woreda output per hectare (OPH) measure in quintals, ranging from low productivity (dark purple) to high productivity (light green).
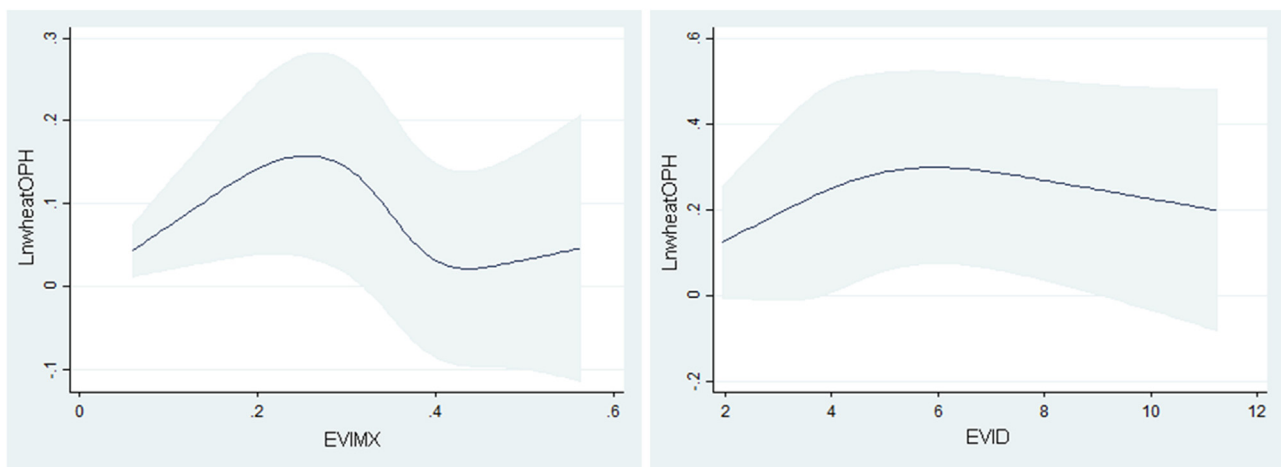
Fig. 3. a,b: Effect of maximum EVI value and area under the declining portion of the EVI curve.
[Left − A] Cubic spline estimation of effects of the maximum Meher season EVI value (EVIMX) on the natural log of wheat output per hectare (LnwheatOPH). [Right − B] Cubic spline estimation of effects of the area under the declining portion of the EVI curve for the Meher growing season (EVID) on the natural log of wheat output per hectare (LnwheatOPH).

were likely not screened out (as described in Section 2.2.4 *Remotely Sensed Data*). Fig. 3b depicts the EVID spline function. For lower values of EVID, a one standard deviation increase (1.79 unit increase in area under the "grain filling" area of EVI) corresponds to a 19.35% increase in productivity (p < 0.01). These increases then taper out at around 50% of the maximum EVID value.

Of the ten spatial variables, five were of the expected sign and statistically significant (*Dist_Addis, Elev, Slope, AgroEco, Pop_Den*), and five were dropped from the In-Sample 2 estimation due to statistical insignificance (*Road_Den, CEC, AGP, CEN_Y & CEN_X*).

As would be expected, the changing nature of the physical terrain impacts productivity. We find that significant increases in elevation (*Elev*) improve output per hectare until approximately 2000 m above sea level (p < 0.05), see Fig. 4a. Mean *kebele* terrain slope (*Slope*) decreases output per hectare until approximately 10° after which yields see no further declines (p < 0.05). The spline fit for slope also indicates increasing yields at the highest angles which is potentially due to the effects of the extensive use of terracing that mitigate problems of steep terrain (Fig. 4b).

The effects of urban primacy (*Dist_Addis*) and population density (*Pop_Den*) are outlined in Fig. 5a,b. The distance to the national capital (*Dist_Addis*) is measured by an increasing Euclidian distance from Addis Ababa (*Dist_Addis*). The spline coefficients correspond to lower agricultural yields until they level out around 300 km (Fig. 5a p < 0.05). Yields increase with *kebele* population density up to around 400 people per square kilometer and then decline beyond this point (Fig. 5b, p < 0.05). However, it is important to note that the median value of population density is 155 people per square kilometer in our sample and the larger values likely reflect fragmented plots in unusually small kebeles.

We also find that the World Bank and Ministry of Agriculture's AGP intervention *kebeles* perform better but not significantly better (p = 0.12) than non-AGP *woredas*. This will be discussed further in Section 4. Agroecological zone fixed effects (*AgroEco*) capture significant differences between growing regions (p < 0.10). We also test the influence of a suite of potential edaphic characteristics (e.g., pH balance, organic material, clay content). Similarly, to Cation-exchange capacity (*CEC*), we find no significant relationship between edaphic properties and wheat output per hectare (p > 0.19, In-Sample 1).

3.2. Yield gap analysis

Yield gap estimates presented here are based on multidimensional clustering that compare *kebeles* with similar climate, weather, and terrain; aggregated to the *woreda* level. The spatial distribution of *woreda* level yield gaps, as determined by climate clusters, can be found in Fig. 6. These yield gaps are the ratio of panel estimates from Eq. (1) and the 90th percentile of the clusters distribution as described in Eq. (3). As such, large negative numbers, for example −80, would indicate at a given *woreda*, on average, is producing 80% less than the 90th percentile *kebele* within the same agroecological cluster. Meanwhile smaller negative numbers, like −25, would indicate that on average, the *woreda* is producing 25% less than the top performing (90th percentile) *kebeles* in their cluster. Regional summaries of gap results from using the cluster based methodology can be found in Table 3.

4. Discussion

Given the small scale of Ethiopia's agriculture and the spatial heterogeneity of it's agroecological zones, explaining approximately 40% of variation in productivity lends statistical support to the further development of this approach. We believe that multidisciplinary approaches, such as this one, can provide insight into

Table 3
Wheat yield gaps by region determined by climate cluster methodology.

Region	Pred OPH	Yield Gap%		
	Median	Median	Min	Max
Amhara	13.96	−51.28	−34.88	−78.44
Oromia	14.04	−47.96	−14.22	−76.54
SNNP	13.85	−49.97	−18.97	−90.38
Tigray	16.80	−39.61	−24.54	−67.82

determinants of agricultural productivity, the effects of management practices, interventions, as well as weather and climate.

4.1. Panel wheat output per hectare results

The Agricultural Growth Project (AGP), administered by the Ethiopian Ministry of Agriculture and funded by the World Bank, aims to increase agricultural productivity and market access for key crops and livestock products in targeted *woredas* (Agriculture and The World Bank, 2011). One of the AGP's two primary objectives is to increase agricultural yields for participating households. Controlling for other determinants in Eq. (1), we see a positive, yet not statistically significant, increase in output per hectare for wheat (p < 0.12). Coefficients indicate that AGP participating *kebeles'* yields are 4.3% higher than non-AGP *kebeles*,[12] compared to a 3.3% increase reported in the AGP baseline paper (IFPRI and EDRI, 2013). This comparison provides some basic support to the predictive ability of our model within the context of other independent analysis.

The AgSS also allows for the control of a variety of input and management relevant variables. We estimate significant gains in productivity from increasing the proportion of land devoted to wheat production. Starting at lower proportions of planted wheat, a one standard deviation increase in the proportion of land planted (+12%) in wheat, increases by 18% output per hectare. Given this finding, the government's newly developed cluster strategy may facilitate significantly increased yields in wheat clusters.[13] The relative importance of this variable should be tempered by its likely endogenous nature, with higher yields encouraging higher emphasis on wheat, and vice versa. That being said, this finding may also point to economies of scale or benefits of specialization for many suitable wheat producing regions, as increased scale likely brings with it lower input and transaction costs, and more capital investment, amongst others. Explanations involving the importance of social capital might also be relevant. Additionally, the stability of regression coefficients across specifications (including specification In-Sample 3 (Table A2 in Appendix A), where *Land-Wheat* is removed) points to this variable's importance and validity for predictive purposes. We find that the self-reported application of improved seeds has a significant influence on wheat yields (p < 0.05). Despite its statistical significance the relative size of their impact remains low. This may be for a variety of reasons. For example, the reported underestimation of improved seed use.[14] (Spielman et al., 2011) is consistent with our finding of high returns from a moderate level of improved seed use. Additionally, there may not be widespread diffusion of those improved seeds most relevant to farmers' challenges (e.g. rust resistant strains). It is advisable to, "Reduc[e] the area currently occupied by susceptible wheat varieties" and "It is highly advisable to release and promote

---

[12] For those *kebeles* participating both in the AgSS survey and AGP interventions.
[13] Personal communications with relevant Ethiopian government officials.
[14] The AgSS questionnaire specifically asks whether or not farmers purchased seed during the crop season and not whether they used improved seed. In many cases farmers either recycle seed or receive improved seed from other farmers, this would suggest larger overall improved seed use.
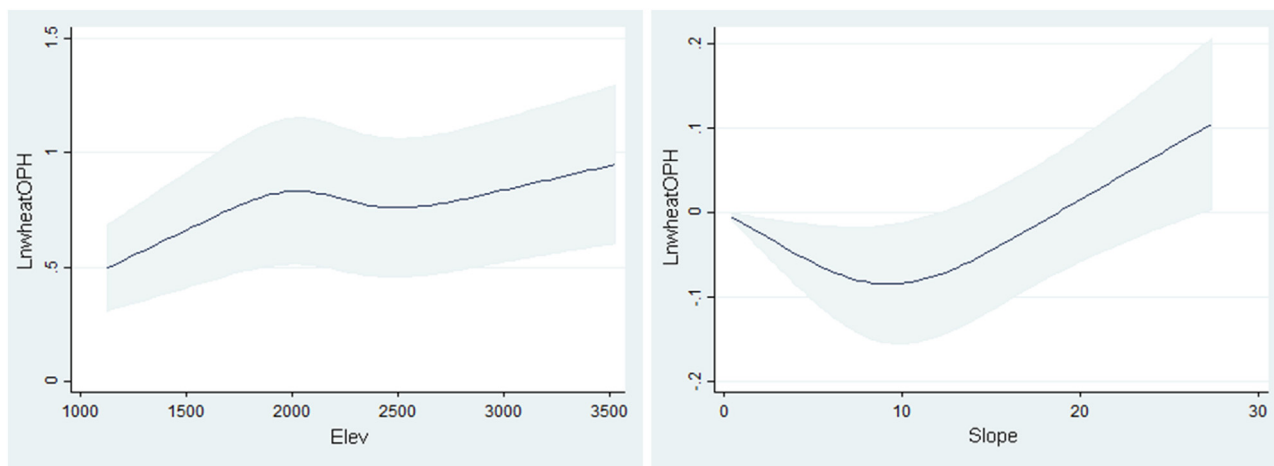
Fig. 4. a,b: Effect of elevation and slope.
[Left − A] Cubic spline estimation of effects of elevation (Elev) on the natural log of wheat output per hectare (LnwheatOPH). [Right − B] Cubic spline estimation of effects of terrain slope (Slope) on the natural log of wheat output per hectare (LnwheatOPH).
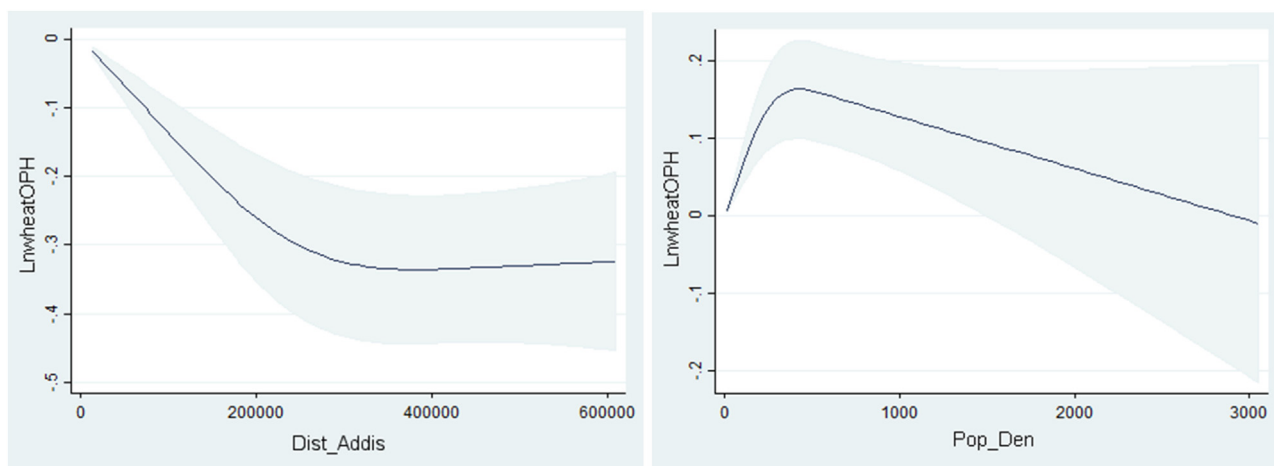
Fig. 5. a,b: Effects of distance to Addis Ababa and population density.
[Left − A] Cubic spline estimation of distance to Addis Ababa (Dist_Addis) on the natural log of wheat output per hectare (LnwheatOPH). [Right − B] Cubic spline estimation of effects of population density (Pop_Den) on the natural log of wheat output per hectare (LnwheatOPH).

varieties that have durable adult plant resistance or have effective race-specific resistance genes in combinations to prevent further evolution and selection of new virulences that lead to boom-and-bust cycles of production" (Singh et al., 2011). This is echoed in the results on damages to agricultural holdings, where we find consistent and significant losses due to pests, rust, flooding and other risks to crops. For instance, for the 2012–2013 period, we find that 11% (3,948/35,844) of farmers and 9.3% (st. dev. = 20%) of all land planted in wheat experienced losses for one of these listed reasons. Interventions, therefore, might emphasize loss prevention methods such as pest and flood control, or the use of fungicide and improved seeds to ward off molds and rust.

As expected, chemical fertilizers increase output per hectare. We find that increasing the application of chemical fertilizer to 95% of all cultivated land dedicated to wheat in a typical *kebele* would increase wheat production per hectare by an additional 3.38%. This speaks to the wide spread diffusion of fertilizers as well as gains that can be obtained through improved inputs and management. Note however that fertilizers, given their current rates of application, have a relatively small impact on obtained yields. This may imply that fertilizers are applied in insufficient quantities, are of

relatively low quality or are not well matched to soil deficits.[15] Although irrigation is likely a critical component of increasing and maintaining high productivity, we are currently unable to estimate its effects at the *kebele* level. This is likely due to the low percentage of households with access to even small-scale irrigation (less than 3% of households), and relatively good growing conditions for the sample period. That being said, findings here still point to a high level of sensitivity to changes in weather and climate. Irrigation, therefore, should and will play a significant role in food security and climate adaptation going forward.

Year to year we see substantial volatility in AgSS measures of agricultural output per hectare. One of the goals of this study was to evaluate whether or not these changes were due to measurement error or driven by the erratic nature of the small-scale rain-fed agricultural systems. In order to tease out potential weather-related effects, we control for a set of potential determinants of inter-

---

[15] To better identify fertilizer needs, Ethiopia's Ministry of Agriculture, working with the Agricultural Transformation Agency (ATA), are currently undertaking a national soil-sampling project (EthioSys) to determine local soil qualities. Fertilizer blending plants have recently been constructed to provide locally appropriate fertilizers.
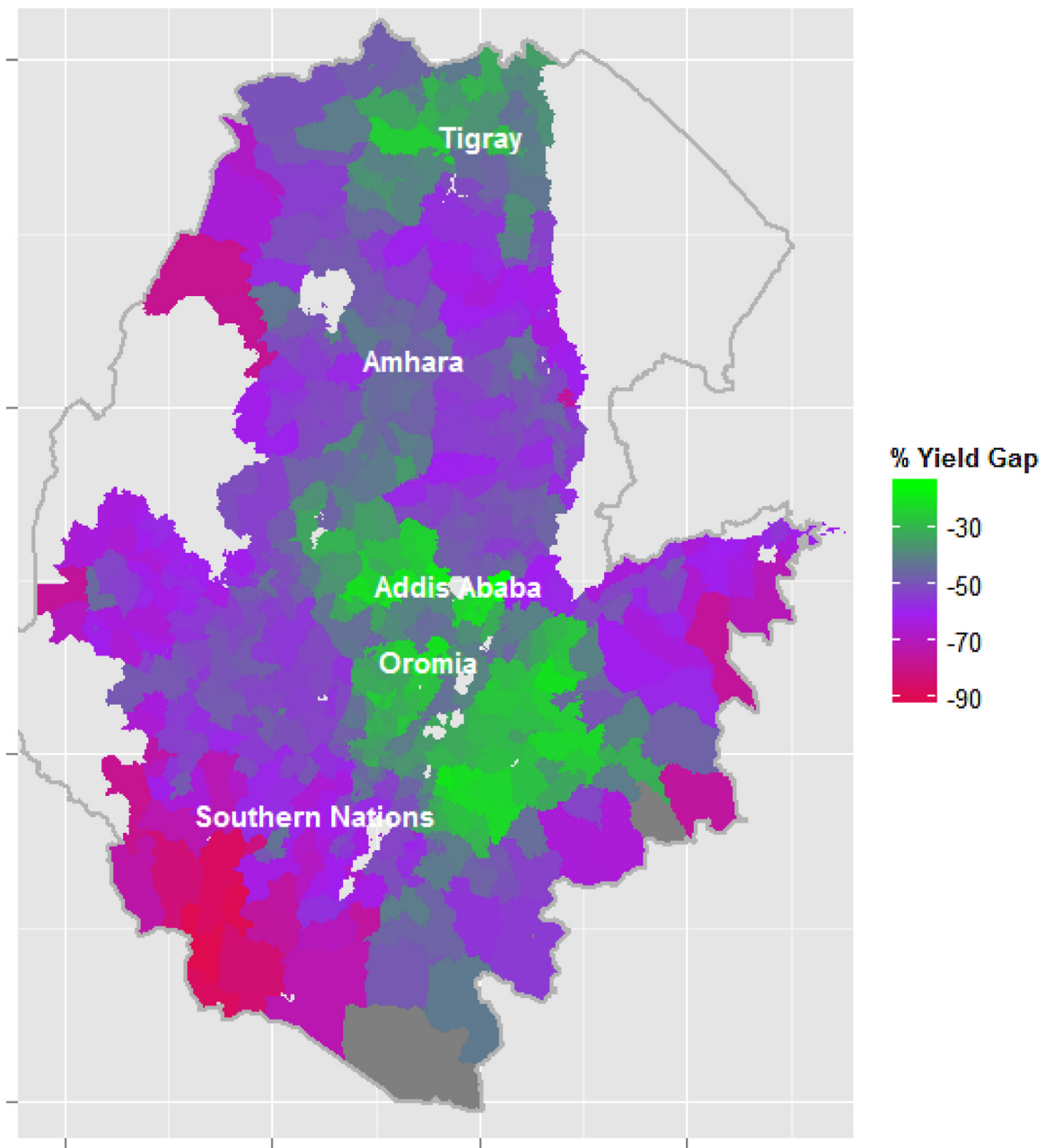
Fig. 6. Spatial distribution of yield gaps as determined by climate clusters.
Woreda-level yield gaps as expressed by the ratio of estimated output per hectare and the 90th percentile kebele within the same agroecological cluster. High yield gaps are shown in red, moderate in purples, and low in light green.

annual variability. These time-variant controls include measures of rainfall, water availability, and plant health observed from satellites. We used measures of the enhanced vegetation index (see Section 2.2.4 for more detail) as a proxy for water availability and crop health at two critical periods of plant growth. The first, the area under the declining portion of the EVI curve (*EVID*), is a valuable measure of plant health and water availability through some of the most critical phases of head development and yield formation. With increases in *EVID*, we see substantial increases in productivity because of favorable conditions for plant growth. Here starting at low levels, a one standard deviation increase (1.79 units) in *EVID* corresponds to a 19.35% increase in yields per hectare. This finding indicates that even in this challenging, small-scale heterogeneous environment, traditional satellite measures of plant health can be applied. It also suggests that despite favorable rains across much of

the country during this study period, there is substantial variation in productivity due to changes in water availability and, therefore, plant health. The slight declines in productivity (and substantial variability around these levels) for high levels of *EVID* may point to the effects of late rains, which increase the likelihood of molds and other diseases late in the growing season. The second EVI index, the maximum annual EVI value, is reached as the maximum levels of chlorophyll and leaf area are reached during the *Meher* growing season. Here pixels with healthy productive plants and high leaf areas will have large maximum values. Looking at Fig. A2, we can see that even wet agricultural areas rarely have maximum EVI values above 0.40. Looking at Fig. 3, we can see that productivity declines rapidly in plots with EVI max values over this level. This is very likely because these pixels contain a higher percentage of tree cover. Further processing of data could reduce the noise asso-

ciated with mixed EVI signatures by better screening out mixed land classes or non-wheat plots. Additionally, in order to avoid challenges on the basis of endogeneity of EVI related variables, we estimate specification In-Sample 4 (Table A2 in Appendix A). Importantly the omission of *EVIMX* and *EVID* points to the stability of other coefficients of interests, as well as the predictive power of EVI variables to the model, with the R-square dropping from 0.38 to 0.27.

We also include time-invariant measures of historical climate and climate variability, terrain characteristics, edaphic properties, and fixed effects indicators such as agroecological zones. Historical climate is proxied with the plant relevant metric of climatic water deficits, which provides a good approximation of historical water availability. We see consistent declines in productivity in areas with historically high water deficits. Because CWD integrates information about precipitation along with soil characteristics and topography, it is a long-term indicator of conditions favorable for plant growth. For instance, sandy soils with steep slopes will likely have high CWD measures and be generally unfavorable for plant growth while flat loamy soils will likely have the opposite effect. As such, CWD and EVI likely capture some of the critical soil (edaphic) and terrain properties that were found insignificant in these regressions, such as CEC and a suite of other properties from AfriSIS data that were dropped due to statistical insignificance (e.g., pH balance, organic matter, clay content, terrain aspect). Additionally, CWD is an indicator of climate expectations of farmers and will likely influence choices such as crop planting type. The inclusion of climatic variability (*CWDSD*) captures some of the effects of climate uncertainty on wheat productivity. Farmers in areas with higher variability in rainfall and relatively unfavorable soil and topographic characteristics (higher *CWDSD*) would be unlikely to make longer-term investments that might boost productivity. Statistically significant fixed effects controls at the zonal and agroecological level speak to the effects of other not explicitly observed regional characteristics such as policy choices and expected growing season length.

Topography, demographics, and distance to key cities also affect productivity. From Fig. 4a we can see that elevation enhances productivity up to a point. The specific relationship between productivity and elevation needs careful attention as the spline function indicates. While both extremely high elevation and increasing terrain slope decreases productivity, this need not be a death knell. This is likely a testament to the Ethiopian people as they terrace and plant in areas not initially suitable for agriculture. Even high elevations do not necessarily imply inhospitable climates and growing conditions (Ethiopia is a parable itself as it is one of the highest elevation crop producing countries in the world). Instead variation in topography and localized conditions in Ethiopia may allow for small pockets of rarified ideal conditions. Beyond topography, the model explores some basic distance and population demographic effects. The model demonstrates that output productivity declines with greater distances from Addis Ababa (Fig. 5a). This finding needs additional research but could be related to relative geographic remoteness and limited access to inputs, investment and therefore capital accumulation. Productivity also varies according to population density, with productivity initially increasing with population density. This may be the result of improved access to inputs, labor, capital, or social capital and declines thereafter as rural landscapes transition to more urban ones.

4.2. Yield gap

Yield gaps are calculated by the ratio of a *kebele*'s wheat yield estimates and the 90th percentile *kebele* operating in similar agroecological conditions. Looking at Table 3 and Fig. 6, we see

substantial gaps between median *woreda*-level performances relative to the 90th percentile *kebeles* in each climate cluster.[16] Gap estimates here should be slightly less than what is typically presented for at least two reasons. First, Y_L is obtained from yields reported by farmers in the AgSS survey, and not the experimental farm or water-limited yields. Second, our methodology aggregates to the *kebele* level, thereby reducing individual outliers. The result is a lower, but more generalizable production gap. It can be generally assumed that the reference technologies are similar across all producers, so closing these gaps are reasonably attainable with appropriate interventions such as improved management or low cost inputs. The estimated gaps provided here are most likely to be closed by providing greater access to improved agro-economic management practices (Nin-Pratt et al., 2011).

At the median, Tigray has the lowest gaps, averaging around 40% less than the top yielding *kebeles* in their climate clusters. This finding speaks to the power of intervention to overcome Tigray's challenging environments; despite having often less than ideal conditions (higher temps, lower rainfall) Tigray's productivity is in line with the other major growing regions. The relative successes in Tigray found here, points to potential power of agricultural interventions to broadly increase yields in Ethiopia, even in the most challenging environments. Next for Oromia we find median gaps of just over 45%. Here the range of yield gaps is quite high, with some *woredas* in Oromia producing between 76 and 14% less (on average) than the best performing *kebeles*. This may be due to the fact that Oromia has the lowest rates of improved seed use (37% compared to 49% on average) and below average use of chemical fertilizers (52% compared to 55% on average). Although these differences might also be driven by inputs not included in this model, capital investment or agricultural extension. Amhara and SNNP median gaps lag slightly behind, with median gaps around 50%. These results most likely reflect the difficult growing conditions in some areas of these regions. We see relatively smaller gaps in the central highlands centered on Addis Ababa and extending into central Oromia (Fig. 6). This is the area typically considered the "wheat belt" that has benefited from some recent localized mechanization interventions.

The yield gap maps produced here can be used to evaluate the efficacy of ongoing interventions at an aggregate scale for the 2011–2013 period. Interventions in top performing *woredas* (green in Fig. 6) should be identified and likely emulated elsewhere. Policymakers might then target *woredas* with intermediate yield gaps, as they narrowly lag top performing areas. Specifically looking at Fig. 6, interventions might focus on *woredas* shown in dark purple allowing them to catch up to the better performing *woredas*. Successful interventions applied in nearby top performing *woredas* will likely address pressing agricultural issues that we identified in the first section of this paper, such as addressing rusts and molds (Singh et al., 2011), improving access to fertilizer and soil inputs, and expanding irrigation in weather sensitive areas.

4.3. Model improvements

Moving forward, a number of improvements could be made to increase the accuracy of the models described here. First and foremost, models could be used with household-level productivity estimates. AgSS aggregation at the *kebele* level, while highly desirable relative to *woreda* or zonal level estimates, obscures key sources of variance that can be observed at only the farm or plot

---

[16] Note that no *woreda* is producing at or above the 90th percentile *kebele*. This is because mean *woreda* productivity is being compared to the distribution of *kebele* productivity. Therefore, as discussed in Section 2.3.4, mean *woreda* performance should be expected to lag behind top performing *kebeles*.

level. With *kebele* level data, the inter-annual noise observed in AgSS crop cut data may be due to causes other than measurement error. For instance, catastrophic losses due to disease or pests on as little as five of the 20 households sampled in a *kebele* could significantly decrease estimates. A more spatially explicit examination of the data may allow us to understand the determinants of successful and unsuccessful years better. In lieu of household data sets, the inclusion of additional AgSS years could help differentiate real changes in productivity from noise.

The current model could also be augmented with additional spatial data. For instance, the 250 m resolution MODIS data used in this study could be downscaled using the 30 m Landsat products. This method would allow for the delineation of individual plot level attributes not currently distinguishable in this report.

The model could also be significantly improved by further collaboration with local and regional experts. This should include further discussions with local agronomists, policymakers, academics, as well as other economists. Additionally, the integration of crop growth models for the estimation of spatially explicit 'potential yields' could be used to further enhance estimated gap models.

5. Conclusion

Using several sets of diverse, but publically available data, we develop a comprehensive model that depicts crop productivity at a variety of administrative levels in Ethiopia. To our knowledge, the methodology of combining survey, climate model values, spatial and remote sensed data in an econometric model for small-scale African agriculture has not been explored. This is somewhat understandable given the inter-disciplinary nature of this research. Combining the economist's typical productivity variables (fertilizer, seed, etc.) with agronomist data (moisture availability, crop cuts, etc.), and geographic information (distance to cities, remotely sensed data, etc.) and other relevant information is a relatively new nexus in this form of research. The results presented here indicate a promising venue for research and policy application.

The model presented effectively incorporates a wide variety of data from both the structural aspects of the model as well as the individual coefficients. Explaining approximately 40% of annual *marginal* productivity variation is promising. In addition, the application of non-linear splines is useful for identifying specific contributions to productivity. Finally, yield gaps and identified determinants of productivity point to specific areas for potential interventions. Importantly we also point to areas where interventions have been successful despite challenging growing conditions. These particular interventions might be identified and replicated where applicable. The ability to identify successful interventions in areas with high interannual variation in temperatures and rainfall might be particularly important in the context of climate change.

Overall, this research has a broad range of potential applications, especially from a public policy perspective. These issues include, but are not limited to, identifying causes of yield determinants, monitoring productivity changes, remotely evaluating larger agricultural interventions, analyzing relative yield potentials, and effective intervention for production shortfalls.

Acknowledgments

Appendix A.

Cumulative Water Deficits (CWD)

A simplified visual representation of CWD or 'deficit' depicted in Fig. A1, demonstrates how temperature (T_{max}, T_{min}) and precipitation (*Precip*) controls the availability of water (Supply) which limits the amount of water moved through evaporation from soils and transpiration through plants (Actual Evapotranspiration, AET). Meanwhile temperatures regulate the amount of water demanded through potential evapotranspiration[17] (Potential Evapotranspiration, PET) (Major, 1967). More complex estimations of CWD include specific soil properties, the effects of terrain, slope, and aspect.
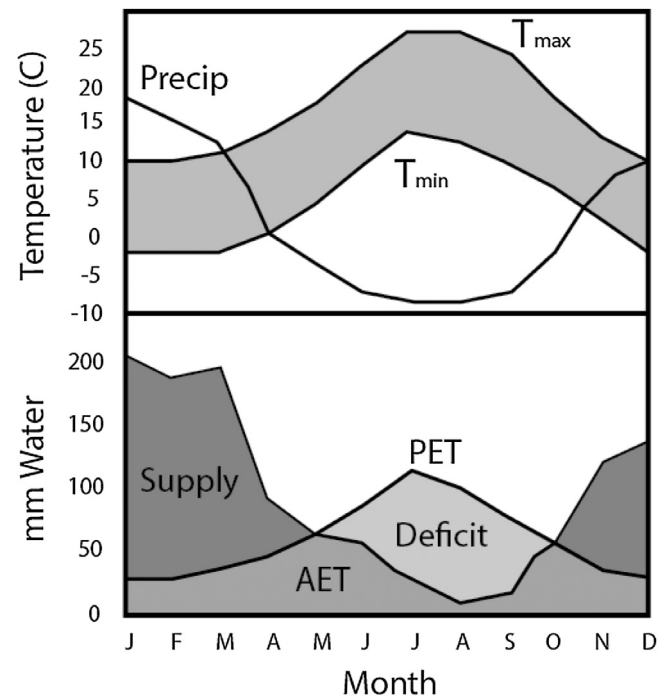

Fig. A1. Climatic Water Deficit by Month.
Climatic water deficit (Deficit) is determined through the interaction of water supply through precipitation (Precip) and evaporation as determined by temperature (T), terrain and edaphic properties. Deficit is the difference between evapotranspiration demanded (Potential Evapotranspiration − PET) and the amount of supplied water actual moved through the system (actual evapotranspiration − AET).

Enhanced Vegetation Index

We removed low quality pixels and used smoothing splines to clean the time series for each of the 44 million pixels for each 16 day period over the three year period. To screen out non-agricultural

---

[17] Pet is the level evapotranspiration by plants and soils when water is a not limiting factor.
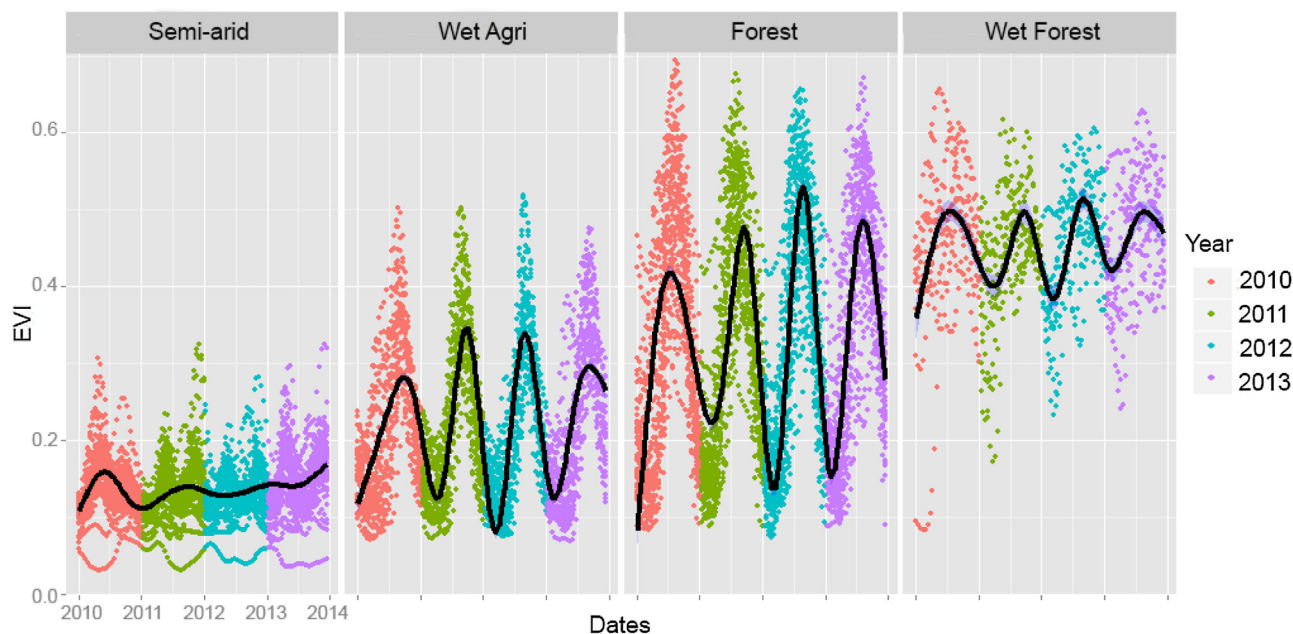
Fig. A2. EVI Time series examples by land cover class.
The time series of the enhanced vegetation index characterizes the phenology of four common land cover classes (Semi-arid, Wet Agriculture, Forest, and Wet Forest) with a smoothing spline (black). Years are shown in colors (Pink-2010, Green-2011, Blue-2012, and Purple-2013).

pixels we used a sample of 200 random points and high-resolution data from Google Earth. For each of these training sites we labeled their land cover and extract their EVI time series. These groups' EVI values then act as a representation of what each land cover "looks like" from space. We can see examples of EVI time series for four key land cover types in Fig. A2. In order to screen out non-agricultural pixels, cells with a correlation of less than 0.5 with wet or dry agriculture were removed from the sample. Given the complexity of the task, we did not attempt to identify individual crop signatures. Therefore, the EVI signals observed for "agriculture" were treated as a proxy for plots planted with wheat. Vegetated land cover types depict a clear seasonal cycle tracing the periodicity of the growing seasons. Looking at the semi-arid land classes, we see low overall values of EVI across the four growing seasons, which reflects the relative lack of vegetation and, therefore, low levels of chlorophyll or "greenness". On the other end of the spectrum are the wet forests of the south, which maintain consistently high levels of EVI across observations over the four-year period. This reflects the lush vegetation and undergrowth of the region, with minimal changes in phenology in any given year. The forest land class shows a clearer seasonality and likely includes deciduous and non-deciduous tree types with less dense undergrowth. Also easily distinguishable is the agricultural signal (Wet Agri), with a rapid green up after planting, a rapid decline during ripening, and the eventual return towards zero as soil is exposed after harvest and prepared for the next growing season.

Table A1
Root mean square error for out-of-sample predictions.

Estimation	RMSE
1	0.33
2	0.29
3	0.23
4	0.38
5	0.24

Out of Sample Performance

We assess the performance of the model *Out-of-Sample* from Eq. (1) using K-fold cross-validation. This procedure splits the data randomly into k partitions. Then for each partition, it fits the specified model using the other k-1 groups and uses the resulting parameters to predict out-of-sample the dependent variable in the unused group. The root mean squared error (RMSE) k-fold results for the final model are as follows:

Full Model Results

We present the results of all model results here. These include 1) "In-Sample 1" an initial specification testing the significance of all variables of interest, 2) "In-Sample 2" a simplified model including all significant variables, 3) "In-Sample 3" drops the possibly endogenous variable *LandWheat*, 4) "In-Sample 4" omits *LandWheat*, all variables using the enhanced vegetation index (EVIMX,

Table A2
Regression on log of output per hectare (standardized coefficients).

	In-Sample 1	In-Sample 2	In-Sample 3	In-Sample 4	Out-of-Sample
LnwheatOPH$_{t-1}$	7.748e-02**	7.992e-02**	9.450e-02**	1.412e-01**	
CWDMN	1.637e-01*	1.643e-01*	1.503e+01+	5.291e-02	8.127e-02
CWDSD	−1.956e-01**	−1.933e-01**	−1.899e-01**	−6.688e-02	−1.351e-01**
Precip	−2.537e-03				
Damage	−9.376e-02**	−9.546e-02**	−1.075e-01**	−1.175e-01**	
EVIMX$^{(S1)}$	7.693e-02*	7.169e-02*	7.578e-02*		8.928e-02**
EVIMX$^{(S2)}$	−1.772e-01*	−1.563e-01*	−1.825e-01*		−2.607e-01**
EVIMX$^{(S3)}$	9.382e-02+	7.945e-02	1.020e-01+		1.557e-01**

Table A2 (*Continued*)

	In-Sample 1	In-Sample 2	In-Sample 3	In-Sample 4	Out-of-Sample
EVID$^{(S1)}$	1.437e-01*	1.769e-01**	1.675e-01*		1.864e-01**
EVID$^{(S2)}$	−5.840e-01*	−6.590e-01*	−7.201e-01*		−8.414e-01**
EVID$^{(S3)}$	4.426e-01+	4.974e-01*	5.561e-01*		6.403e-01**
LandWheat$^{(S1)}$	1.738e-01*	1.693e-01*			
LandWheat$^{(S2)}$	−9.231e-02	−8.717e-02			
ImprSeed$^{(S1)}$	−1.621e-03	−2.206e-03	4.298e-02	6.479e-02	
ImprSeed$^{(S2)}$	9.093e-02	8.742e-02	−3.662e-02	−7.226e-02	
ImprSeed$^{(S3)}$	−1.306e-01	−1.262e-01	−4.291e-02	−2.809e-02	
ChemFert	3.340e-02+	3.332e-02+	4.344e-02*	5.219e-02**	
Irrigation	4.658e-04				
Elev$^{(S1)}$	8.770e-01**	8.629e-01**	8.440e-01*	2.389e-01	9.424e-01**
Elev$^{(S2)}$	−1.233e+00**	−1.212e+00**	−1.100e+00*	−2.918e-01	−1.100e+00**
Elev$^{(S3)}$	6.119e-01**	6.008e-01**	5.049e-01*	1.359e-01	4.468e-01*
Slope$^{(S1)}$	−6.786e-02+	−6.376e-02+	−4.172e-02	−6.224e-02+	−2.194e-02
Slope$^{(S2)}$	9.570e-02**	8.133e-02*	7.129e-02*	4.999e-02+	6.498e-02*
CEC	−2.220e-02				
AgroEco	–	–	–		–
Dist_Addis$^{(S1)}$	−1.806e-01**	−1.549e-01*	−1.917e-01**	−1.619e-01**	−1.653e-01**
Dist_Addis$^{(S2)}$	1.857e-01*	1.704e-01*	2.154e-01**	1.247e-01**	1.744e-01**
Road_Den$^{(S1)}$	4.555e-02				
Road_Den$^{(S2)}$	−7.157e-02				
Pop_Den$^{(S1)}$	2.055e-01*	2.119e-01*	2.016e-01*	2.864e-01**	
Pop_Den$^{(S2)}$	−2.641e-01**	−2.738e-01**	−2.660e-01**	−3.652e-01**	
AGP	4.318e-02	4.267e-02	3.606e-02	4.700e-02+	5.241e-02*
Cen_Y	−6.190e-03				
Cen_X	−9.940e-02				
YEAR	2.037e-03				
Z_Code	–	–	–		–
Constant	2.643e+00**	2.593e+00**	2.624e+00**	2.706e+00**	2.740e+00**
R^2	0.39	0.38	0.37	0.27	0.24
N	1726	1726	1726	1728	2907

+90%, *95%, **99% level of significance.

EVID), and all fixed effects, and 5) "Out-of-Sample" omits all variables related to the AgSS survey, in order to allow for out-of-sample yield estimates. In-Sample 1–2 and Out-of-Sample are identical to the results presented in Table 2 in the text.

References

AgSS, 2012. Central Statistical Authority of Ethiopia (CSA). Agricultural Sample Survey (AgSS 2011–2012) [WWW Document]. URL www.csa.gov.et/.

AgSS, 2013. Central Statistical Authority of Ethiopia (CSA). Agricultural Sample Survey (AgSS 2012–2013) [WWW Document]. URL www.csa.gov.et/.

AgSS, 2014. Central Statistical Authority of Ethiopia (CSA). Agricultural Sample Survey (AgSS 2013–2014) [WWW Document]. URL www.csa.gov.et/.

Agriculture, The World Bank, 2011. Ethiopia − Agricultural growth project (AGP). Project information document (PID).

Ahmad, I., Ghafoor, A., Iftikhar, Muhammad, Bhatti, M.I., Akhtar, I.H., Ibrahim, M., Rehman, O., 2014. Satellite remote sensing and GIS based crops forecasting & estimation system in Pakistan, islamabad. In: Pakistan: Space Applications and Research Complex. Pakistan Space and Upper Atmosphere Research Commission.

Alemu, T., Emana, B., Legesse, G., 2014. Smallholder wheat production efficiency in selected agroecological zones of Ethiopia: a parametric approach. J. Econ. Sustain. Dev. 5.

Altieri, M., Koohafkan, P., 2008. Enduring Farms: Climate Change, Smallholders and Traditional Farming Communities (Penang, Malaysia).

Cliff, A., Ord, J., 1981. Spatial Processes: Models and Applications. Pion, London.

Didan, K., Huete, A., 2006. MODIS Vegetation Index Product Series Collection 5 Change Summary.

Durrleman, S., Simon, R., 1989. Flexible regression models with cubic splines. Stat. Med. 8, 551–561.

FAO, 2013. Crop Water Information: Wheat. FAO (WWW Document).

FDRE, 2013. Ethiopia's Climate Resilient Green Economy: CLIMATE RESILIENT STRATEGY AGRICULTURE. FDRE, Addis Ababa, Ethiopia.

Ferenz, C., Bognar, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timar, G., Molna Pasztor, G.R.S., Steinbach, P., Szekely, B., Ferenc, O., Ferencz-Arkos, I., 2004. Crop yield estimation by satellite remote sensing. Int. J. Remote Sens. 25, 4113–4149.

Fontana, A., Potgieter, B., Apan, A., 2005. Assessing the relationship between shire winter crop yield and multi-temporal MODIS NDVI and EVI images. In: The National Biennial Conference of the Spatial Sciences Institute, Proceedings of SSC 2005 Spatial Intelligence, Innovation and Praxis: The National Biennial Conference of the Spatial Sciences Institute, Melbourne.

GSFC, 2014. Tropical Rainfall Measurement Mission Project (TRMM), Daily TRMM and Others Rainfall Estimate (3B42 V7 derived), version 7, [WWW Document]. URL http://disc.sci.gsfc.nasa.gov/recipes/?q=datacollection/TRMM_3B42_daily.007/description

Gelaw, F., Bezabih, E., 2009. Analysis of technical efficiency of wheat production: a study in Machakel Woreda, Ethiopia. Ethiop. J. Agric. Econ. 7, 1–33.

Huber, P., 1967. The behavior of maximum likelihood estimates under nonstandard conditions. Proc. Fifth Berkeley Symp. Math. Stat. Probab. 1, 221–233.

IFPRI, EDRI, 2013. Agricultural Growth Program (agp) of Ethiopia — Baseline Report 2011. IFPRI, EDRI, Addis Ababa.

IFPRI, 2013. Agricultural Growth Program (agp) of Ethiopia − Baseline Report 2011. ESSP II, Washington DC.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled SRTM for the globe Version 4. available from the CGIAR-CSI SRTM 90 m Database.

Jones, P.G., Thornton, P.K., 2003. The potential impacts of climate change on maize production in Africa and Latin America in 2055. Glob. Environ. Chang. 13, 51–59.

Lachenbruch, P., Mickey, M., 1968. Estimation of error rates in discriminant analysis. Technometrics 10, 1–12.

Leenaars, J.G.B.G.B., 2013. Africa Soil Profiles Database Version 1.1. A Compilation of Geo-referenced and Standardized Legacy Soil Profile Data for Sub Saharan Africa (with Dataset). ISRIC Report 2013/03 (Wageningen, Netherlands).

Licker, R., Johnston, M., Foley, J., Barford, C., Kucharik, C., Monfreda, C., Ramankutty, N., 2011. Mind the gap: how do climate and agricultural management explain the yield gap of croplands around the world? Glob. Ecol. Biogeogr. 19, 769–782.

Liu, J., Liu, M., Tian, H., Zhuang, D., Zhang, Z., Zhang, W., Tang, X., Deng, X., 2005. Spatial and temporal patterns of China's cropland during 1990–2000: an analysis based on Landsat TM data. Remote Sens. Environ. 98, 442–456, http://dx.doi.org/10.1016/j.rse.2005.08.012.

Lobell, D.B.J., Ortiz-Monasterio, J., Asner, G., Naylor, R., Falcon, W., 2005. WHEAT—combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. Agron. J. 97, 241–249.

Major, J., 1967. Potential evapotranspiration and plant distribution in western states with emphasis on California. AMER ASSOC Adv. ScI. Publ. 86.

Mann, M., Warner, J., 2015. Ethiopian Wheat Yield and Yield Gap Estimation: A Small Area Integrated Data Approach. Research for Ethiopia's Agricultural Policy, Addis Ababa, Ethiopia.

Minot, N., Warner, J., Lemma, S., Abate, G., Rashid, S., 2015. The Wheat Supply Chain in Ethiopia: Patterns, Trends, and Policy Options (Addis Ababa, Ethiopia).

Moran, M.S., Inoue, Y., Barnes, E.M., 1997. Opportunities and limitations for image-based remote sensing in precision crop management. Remote Sens. Environ. 61, 319–346, http://dx.doi.org/10.1016/S0034-4257(97)00045-X.

Nin-Pratt, A., Johnson, M., Magalhaes, E., You, L., Diao, X., Chamberlin, J., 2011. Yield Gaps and Potential Agricultural Growth in West and Central Africa (Washington D.C.).

Population Census Commission, 2008. Summary and Statistical Report of the 2007 Population and Housing Census: Population Size by Sex and Age (Addis Ababa, Ethiopia).

Prasad, A.K., Chai, L., Singh, R.P., Kafatos, M., 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Obs. Geoinf. 8, 26–33, http://dx.doi.org/10.1016/j.jag.2005.06.002.

Randall, L., Bruce, S., Nikolova, S., Lawson, K., 2011. A Review of the Use of Remote Sensing for Crop Forecasting in Australia (Canberra: Australia).

Serrano, L., Filella, I., Peñuelas, J., 2000. Remote Sensing of Biomass and Yield of Winter Wheat Under Different Nitrogen Supplies.

Singh, R.P., Hodson, D.P., Huerta-Espino, J., Jin, Y., Bhavani, S., Njau, P., Herrera-Foessel, S., Singh, P.K., Singh, S., Govindan, V., 2011. The emergence of ug99 races of the stem rust fungus is a threat to world wheat production. Annu. Rev. Phytopathol. 49, 465–481.

Sosa-Escudero, W., Bera, A.K., 2008. Tests for unbalanced error-components models under local misspecification. Stata J. 8, 68–78.

Spielman, D., Mekonnen, D., Alemu, D., 2011. Seed, Fertilizer, and Agricultural Extension in Ethiopia, in Food and Agriculture in Ethiopia-Progress and Policy Challenges. University of Pennsylvania Press, Philadelphia, PA, USA.

Swinton, S.M., Lowenberg-Deboer, J., 2001. Global adoption of precision agriculture technologies: who, when and why. Proceedings of the 3rd European Conference on Precision Agriculture, 557–562.

Taffesse, A., Dorosh, P., Asrat, S., 2011. Crop Production in Ethiopia: Regional Patterns and Trends ESSP II Working Paper No. 0016 (Addis Ababa, Ethiopia).

The World Bank, 2013. World Development Indicators: Country Level Data for Ethiopia [WWW Document]. The World Bank http://data.worldbank.org/country/ethiopia.

UN-OCHA, 2014. Humanitarian Response COD-FOD Registry [WWW Document]. UN-OCHA (URL) http://www.humanitarianresponse.info/operations/ethiopia/dataset/ethiopia-admin-level-1-boundaries-admin-level-2-boundaries-admin-level-3 (accessed 10.9.14.).

WLRC, 2013. Ethio GIS 2. WLRC.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica1 48, 817–830.

Willmott, C.J., Matsuura, K., 2001. Terrestrial Water Budget Data Archive: Monthly Time Series (1950 − 1999) Version 1.02.

Willmott, C.J., 1977. WATBUG: a FORTRAN IV algorithm for calculating the climatic water budget. Publ. Climatol. 30, 1–55.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge, MA.

Yami, M., Solomon, T., Begna, B., Fufa, F., Alemu, T., Alemu, D., 2013. Source of technical inefficiency of smallholder wheat farmers in selected waterlogged areas of Ethiopia: a translog production function approach. Afr. J. Agric. Res. 8, 3930–3940.