Climate Risk Management 35 (2022) 100396

Contents lists available at ScienceDirect

Climate Risk Management

journal homepage: www.elsevier.com/locate/crm

Using real-time mobile phone data to characterize the relationships between small-scale farmers' planting dates and socio-environmental factors

Natasha Krell [a], Frank Davenport [b], Laura Harrison [b], William Turner [a], Seth Peterson [a,b], Shraddhanand Shukla [b], Jessica Marter-Kenyon [c], Greg Husak [b], Tom Evans [d], Kelly Caylor [a,e]

[a] Department of Geography, University of California, Santa Barbara, United States
[b] Climate Hazards Center, University of California, Santa Barbara, United States
[c] College of Agricultural & Environmental Sciences, University of Georgia, United States
[d] School of Geography, Development and Environment, University of Arizona, United States
[e] Earth Research Institute, University of California, Santa Barbara, United States

ARTICLE INFO

Keywords:
Planting dates
Start of season
Small-scale food producers
Central Kenya
Mobile phone data
Random forests

ABSTRACT

Accurate and operational indicators of the start of growing season (SOS) are critical for crop modeling, famine early warning, and agricultural management in the developing world. Erroneous SOS estimates–late, or early, relative to actual planting dates–can lead to inaccurate crop production and food-availability forecasts. Adapting rainfed agriculture to climate change requires improved harmonization of planting with the onset of rains, and the rising ubiquity of mobile phones in east Africa enables real-time monitoring of this important agricultural decision. We investigate whether antecedent agro-meteorological variables and household-level attributes can be used to predict planting dates of small-scale maize producers in central Kenya. Using random forest models, we compare remote estimates of SOS with field-level survey data of actual planting dates. We compare three years of planting dates (2016–2018) for two rainy seasons (the October-to-December short rains, and the March-to-May long rains) gathered from weekly Short Message Service (SMS) mobile phone surveys. In situ data are compared to SOS from the Water Requirement Satisfaction Index (SOS_{WRSI}) and other agro-meteorological variables from Earth observation (EO) datasets (rainfall, NDVI, and evaporative demand). The majority of farmers planted within 20 days of the SOS_{WRSI} from 2016 to 2018. In the 2016 long rains season, many farmers reported planting late, which corresponds to drought conditions. We find that models relying solely on EO variables perform as well as models using both socio-economic and EO variables. The predictive accuracy of EO variables appears to be insensitive to differences in reference periods that were tested for deriving EO anomalies (1, 3, 5, or 10 years). As such, it would appear that farmers are either responding to short-term weather conditions (e.g., intra-seasonal variability), or longer trends than were included in this study (e.g., 25–30 years), when planting. The methodologies used in this study, weekly SMS surveys, provide an operational means for estimating farmer behaviors–information which is traditionally difficult and costly to collect.

https://doi.org/10.1016/j.crm.2022.100396
Received 5 April 2021; Received in revised form 31 October 2021; Accepted 7 January 2022
Available online 12 January 2022

1. Introduction

Hundreds of millions of farmers around the world depend on rainfed agriculture for their primary source of livelihood (Dixon et al., 2001; Rockstrom et al., 2007). However, increasing climate uncertainty and unfavorable weather conditions challenge agricultural production and can lead to adverse food-security outcomes. Specifically, delayed onset of rains can cause smallholder farmers to plant crops outside the optimal window and lead to reduced yields or total crop failure (Abrahão and Costa, 2018; Jain et al., 2015). In east Africa, specifically Kenya, frequent droughts and lower average seasonal rainfall during the main agricultural growing season have characterized recent decades (Funk et al., 2019; Liebmann et al., 1996; FEWS NET, 2019), and this raises concern for a drying trend becoming a "new normal." Rainfall extremes and flood risks, such as during the very wet short rains in 2019, also highlight the large variability that farmers in this region face (Wainwright et al., 2021; FEWS NET, 2019). Between April 2011 and February 2020, the majority of Kenya was reported to be in a "Crisis" phase of food insecurity (Shukla et al., 2021). The onset date of the rainy season is a key driver of food security and livelihood for small-scale farmers who depend largely on rainfed agriculture.

The timing of planting (sowing) is one of the most-critical decisions an agriculturalist will make in rainfed farming conditions and is highly impacted by the rainy season onset. Studies have shown the potential for improved yields when the date of planting is timed with rainfall onset (Sultan et al., 2005). If a farmer plants too early the seed may fail to germinate due to insufficient soil moisture. If a farmer plants too late, there may be insufficient rainfall during the later parts of the growing season, which are crucial for crop growth for staple crops such as maize. In east Africa in particular, the highly variable nature of rainfall and general lack of access to irrigation places extreme importance on planting dates.

Understanding the relationship between planting decisions and social and environmental factors requires estimations of sowing dates. These sowing dates can either be observed or estimated. In general, there is a lack of observed sowing date information, especially for large-scale analyses in sub-Saharan Africa (SSA). Even at the village level, data on sowing dates is generally lacking because it is difficult and expensive to collect. Traditional methods of collecting planting date information include household surveys or, more recently, remotely-sensed estimations (Zhang, 2020). However, the decision to plant is one that is largely made in real-time by farmers, and as such, is susceptible to fast changes in the environment or weather forecasts. Furthermore, the planting date recalled at a later date can prove inaccurate due to faulty recall. Thus, while household surveys can capture broad planting patterns, such as what month they occurred, planting dates should also be ascertained by real-time data collection, such as through mobile phones. Mobile phones offer the potential to collect data in real-time with high frequency, which are otherwise difficult to capture, and thus address data gaps in food security and crop modeling studies (Giroux et al., 2019). Monitoring smallholder farmers' grain storage is one example where mobile phone data collection has elucidated the temporal dynamics of intra-seasonal food security (Waldman et al., 2020).

Planting patterns throughout the world are generally tied to rainfall climatology and the general tenet in the literate states that planting in tropical rainfed systems occurs at the start of the rainy season (Bondeau et al., 2007; Bussmann et al., 2016; Fisher et al., 2015; Laux et al., 2008; Laux et al., 2010). However, while planting dates at the start of the season do maximize available water, these planting dates may be unrealistic at the field level (Bussmann et al., 2016; Laux et al., 2010; Waha et al., 2012). Srivastava et al. (2016) state that the assumed relationship between climate and planting can mask other factors and perhaps more true drivers behind sowing decisions.

Earth observation (EO) data, such as evaporative demand, rainfall, and vegetation greenness change as a result of rainy season onset and therefore correlate with farmers' planting dates. However, the decision of when to plant is complex and driven by multiple factors besides spatio-temporal variability in soil water or rainfall. For example, farm-level biophysical factors (e.g., soil type, soil texture, and topography) interact with rainfall patterns to produce varied soil conditions, which then influence the timing of land preparation and planting (Matlon, 1980). Further, both scientific evidence and common sense indicate that farmer-specific socio-cultural, demographic, economic, technological, and logistical factors interact to influence planting date (Zhang, 2020). Although we have a limited understanding of the complex human drivers (broadly writ) of planting dates, the existing literature does reveal several relationships of interest.

Household attributes related to socio-economic status (such as education, income, and access to various forms of capital) can influence farmers' perceptions of the right time to plant as well as their ability to execute planting-related decisions. Studies in Ethiopia have found that farmers' ability to adjust their planting dates is significantly impacted by income and formal education (Belay et al., 2017; Negash et al., 2011), which provide farmers with the informational and financial capacity they need to adapt to variable weather conditions. Differences in knowledge and information can also lead to variations in crop planting date. Studies show that farmers in Ethiopia and Swaziland, for example, are more likely to alter sowing dates if they have access to extension services and climate information that help them plan for the start of season (Belay et al., 2017; Shongwe et al., 2014; Negash et al., 2011).

Knowledge-related factors intersect with age and other socioeconomic and demographic factors, including gender. For example, late planting of cotton in Cameroon has been linked, in part, to lack of extension for young farmers (Deveze and Halley, 2005). Studies also show that farmers' ability to adapt to climate change through adjustments in planting dates is significantly impacted by gender (Belay et al., 2017), as well as age and the size of social networks (Shongwe et al., 2014). These results are consistent with what we know about agricultural technology adoption more broadly (Van den Berg, 2013; Jain et al., 2015; Krell et al., 2020).

Finally, planting dates are influenced by the amount of land that needs to be planted and how many people are available to assist. Family size can alter sowing dates even within areas where there is a similar time of rain onset because farmers with larger families can plant earlier, and with more flexibility in response to environmental cues due to greater availability of labor (Sacks et al., 2010; Srivastava et al., 2016; Marie et al., 2020). In the Gondar Zuria District of northwestern Ethiopia, Marie et al. (2020) found that both farm and family size were positively correlated with the adoption of early and late planting in male-headed households, which the

Table 1
Advantages and disadvantages of survey type methodologies employed in the study.

Survey Type	Advantages	Disadvantages
SMS	Real-time capture of social-environmental issues Actionable information Relative low cost	Network constraints Respondent fatigue Data mining challenges with missing observations.
Household survey	Higher question volume Ability to probe meaning and intent behind response. Completeness	High cost per observation Time intensive Recall error

authors attribute to improving the resources available to households for implementing new agricultural technologies and practices. However, this is not necessarily a linear relationship: small farm size is often correlated with lower wealth and asset ownership (Harris and Orr, 2014), which may prohibit these farmers from accessing technology, information, and capital that facilitate their ability to meet desired planting dates. On the other hand, farmers with large plots of land may find themselves unable to plant "on time" if they lack adequate machinery or labor (whether paid, family, or animal).

This paper investigates the relationship between household-level variables, EO datasets, and planting decisions among small-scale maize producers in central Kenya. As an applied paper, we investigate the feasibility of using EO datasets to predict planting dates; however, we consider the effect of both social (household-level) and environmental factors on farmer-reported planting dates. Our primary motivation for exploring biophysical variables via EO datasets is that they are already collected with wide temporal and spatial resolutions, and are available in near-real time. In contrast, socioeconomic variables are generally limited in scope, temporal resolution, and in the case of household surveys, are published long after they are collected. The household-level variables chosen are intended to provide a baseline snapshot of the farmers' social and demographic context. Our study uses two methods of planting dates data collection–real-time mobile-phone based surveys and household surveys–as well as household demographic variables and EO datasets to answer the following questions:

1. Can antecedent agro-meteorological variables related to the start of season and household attributes be used to accurately predict the planting dates reported by small-scale maize producers?
2. How much, if any, does prediction improve when we include household variables compared to only EO datasets?
3. Which physical dataset best predicts planting dates?

We used a simple rainfall-accounting methodology–described by AGRHYMET (1996) and implemented operationally in the Water Requirement Satisfaction Index (WRSI)–to define the onset of rains, herein referred to as SOS_{WRSI}. We then subtract the farmers' planting date from the SOS_{WRSI} to get a measure of difference between the planting date and SOS. With this planting difference, we run multiple random forest models to determine which social and environmental variables, and for which farmer geographic locations, planting decisions align best. We find that farmers' planting dates co-vary with the SOS, as the majority of farmers plant within twenty days of the SOS. We investigate which EO datasets and household variables predict planting decisions and find strong predictive relationships for EO data. Lastly, changing the reference period used to define anomalies for EO data results in small and varying differences in predictive accuracy. Overall, we show the utility of using EO data when in-field planting dates are unavailable.

2. Study site

The study was conducted in a semi-arid region in central Kenya, on the western slopes of Mount Kenya. We surveyed small-scale farmers in communities in the Kenyan highlands, specifically in Laikipia, Meru, and Nyeri counties. Rainfall patterns in the study site vary, with higher-elevation areas receiving as much as 1000 mm annual rainfall, and semi-arid lowlands receiving less than 500 mm (Gichangi et al., 2015; Krell et al., 2021). Annual rainfall is bimodal, with wet seasons extending roughly from March to May or early June (long rains), and then again from October to December (short rains). These two wet seasons are separated by two short, dry seasons from July to September, and January to February.

Agricultural production in Kenya is largely rainfed. However, some of the households surveyed as part of this research participate in local water resource governance groups called Community Water Projects (CWPs). Households that have CWP membership receive varying amounts of water for irrigation and domestic use from a network of pipes that distribute water from Mount Kenya (Waldman et al., 2019). Members interact on a weekly to monthly basis in meetings and maintain irrigation infrastructures. To become a member, farmers typically pay a fee to join and to connect to piped networks serving irrigation water and monthly maintenance fees.

Maize is a staple crop grown in the region, and other crops grown at the household level include cowpeas, beans, snow peas, sunflowers, a variety of fruit trees, cabbage, and leafy green vegetables. Farmers have opportunities to sell their crops at local markets in the nearby towns such as Nanyuki and Timau. In addition to small-scale agricultural production, the local economy is also fueled by industrial agriculture of flower farms, canola oil, and other cash-crop farms. Overall, the region is a heterogeneous environment in which multiple commercial producers compete for surface water and groundwater access, which, along with increased population growth, has led to increased demand and decreased water flow over the past 60 years (Ngigi et al., 2007).

Table 2
Description of environmental datasets.

Variable	Product	Spatial Resolution	Temporal Extent	Measure
Rainfall	CHIRPS[a]	0.1° (~10 km)	1981-Present	1,3,5, and 10 year rainfall anomalies during the month of planting
Evaporative Demand	Reference ET (ET0) Monitoring Dataset[b]	0.125° (~12 km)	1981-Present	1,3,5, and 10 year ET0 anomalies during the month of planting
NDVI MAX	NASA MODIS TERRA[c]	0.08° (~8 km)	2001-Present	1,3,5, and 10 year NDVI MAX anomalies during the month of planting

[a]Climate Hazards Center InfraRed Precipitation with Station data.
[b]Uses MERRA-2 atmospheric reanalysis.
[c]Normalized Difference Vegetation.

Table 3
Description of household survey variables used in random forest models and metrics used to assign planting decisions.

Variable	Description	Data Type
County	County of household location.	Categorical
Community Water Project (CWP)	Does the household participate in Community Water Project (CWP). I.e., does the household have access to irrigation.	Binary
Farm size (hectares)	Size of farm plot in hectares (ha).	Continuous
Farmer group	Is the farmer engaged in a farmer group.	Binary
Agricultural cooperative member	Is the farmer engaged in an agricultural cooperative.	Binary
Contact with extension	Has the farmer been in contact with an agricultural extension agent in the past year?	Binary
Total Livestock Units (TLU)	Index of number of livestock owned by household (Jahnke, 1982).	Continuous
Household size (Number of people)	Size of household in number of persons.	Continuous
Education (Highest level attained)	Highest level of education completed by head of household.	Categorical
Planting date	The planting date from either the SMS or household survey.	Continuous
SOS_{WRSI} dekad	Dekad of SOS using AGRHYMET rainfall accounting	Continuous
Planting dekad	Dekad of week of planting.	Continuous
Community average planting dekad	Average planting dekad for community in which the farmer is located.	Continuous
Planting decision	Difference between the Planting dekad and SOS_{WRSI} dekad. Provides a metric of planting decision in which negative numbers indicate early planting relative to the SOS_{WRSI}.	Continuous

3. Data

3.1. Farmer-reported planting dates

We collected farmer responses on planting dates by two means: household surveys (HHS) and Short Message Service (SMS) surveys between 2016 and 2018. We summarize the advantages and disadvantages of the two available sources of planting dates in Table 1. The farmers selected for the study are smallholder farmers living in Laikipia, Meru, and Nyeri counties in central Kenya. We selected the households as part of a five-year multi-institutional research project conducted in communities surrounding Nanyuki, Kenya (Lopus et al., 2017; McCord et al., 2017). The project assessed agricultural decision-making of smallholder farmers enrolled in irrigated community water projects (CWPs), and households without access to irrigation living in the same communities (non-CWP farmers). We selected either heads of household or spouses of the heads of household as survey respondents.

The week of farmers' planting was collected via SMS during two growing seasons. Farmers responded to the question, "Have you planted maize in the last 7 days," and responses were uploaded automatically through an API (textit.in). We converted the week prior to the farmers' response into dekads, which correspond to the first, second, or third 10- or 11-day period of the month, to estimate a planting dekad (WMO, 1992). Therefore, we use an estimate of the week of farmers' planting in dekads, which we refer to as planting dekad (see Table 3). Further details on SMS survey sampling can be found in Appendix A.

We conducted two household surveys in 2017 and 2018, which provided information about farmers' planting dates during the short rains 2016 and 2017 seasons and the long rains in 2018. Importantly, the household surveys captured recalled estimates of farmers planting. For example, we asked about farmers' planting decisions in the 2016 short rains (October to December) during the household survey conducted between March and April 2017. Therefore, there was an approximately 6 month gap between the planting date and the recorded planting date. The only season that was captured in the household survey with less than a six-month gap was the 2018 long rains season. We surveyed farmers between June and July in 2018 who reported their week of planting for the 2018 long rains (approximately 4-month gap or less). Further details about the household survey sampling can be found in Appendix B. The number of farmers who provided planting date information through either survey type is shown in Table 6. The total number of unique farmers that responded to either survey type is 659.

Table 4
Summary statistics of categorical demographic variables.

Variable	Index	*n*	Proportion
County	Laikipia	326	0.50
	Meru	145	0.22
	Nyeri	142	0.22
	No Response	45	0.07
Community Water Project (CWP)	Yes	487	0.74
	No	171	0.26
Farmer group member	Yes	254	0.39
	No	360	0.55
	No Response	44	0.07
Agriculture cooperative member	Yes	83	0.13
	No	531	0.81
	No Response	44	0.07
Contact with extension	Yes	337	0.51
	No	277	0.42
	No Response	44	0.07
Education (Highest level attained)	None	35	0.05
	Some primary	67	0.10
	Completed primary	145	0.22
	Some secondary	81	0.12
	Completed secondary	143	0.22
	Some post-secondary	34	0.05
	Completed post-secondary	54	0.08
	Vocational Training	22	0.03
	Unknown	21	0.03
	No Response	56	0.09

For both the SMS and HHS, GPS coordinates were collected at the households' home. The structure at which the geographic coordinates were collected were, on average, less than 100 meters away from the field where the planting occurred. Therefore, the GPS location and the field that was planted were, in almost all cases, co-located, and thus, the GPS coordinate closely reflects the farmers' field. For future studies that use EO variables with a spatial resolution on the order of 10 m to 500 m, better care would be needed to use GPS coordinates that represent the actual field. For the purpose of this study, where the spatial resolution of the EO variables is between ~8 to 12 km (see Table 2), the GPS location at the farmers' home is sufficient.

3.2. Start of Season (SOS)

In this study, we focus on a metric for determining the SOS using a threshold amount and distribution of rainfall received in three consecutive dekads, as defined by the Agriculture-Hydrology-Meteorology Regional Center in Niger (AGRHYMET, 1996). This metric was selected because it is the primary methodology implemented in the Water Requirement Satisfaction Index—a simplified crop-weather analysis model commonly used by used by the Famine Early Warning Systems Network (FEWS NET) for monitoring ongoing growing seasons, as well as a proxy for crop-yield information at the conclusion of seasons. This methodology indicates that a SOS is established when there is at least 25 mm of rainfall in an initial dekad, followed by a total of at least 20 mm of rainfall in the following two consecutive dekads. Once these rainfall metrics are met, the SOS is defined as the first dekad in that 3-dekad series, and will hereby be referred to as SOS_{WRSI}. The input used for rainfall is the Climate Hazards Center InfraRed rainfall with Station data (CHIRPS) rainfall product. CHIRPS is aggregated from daily to dekadal time steps, and resampled from its native 0.05-degree spatial resolution to 0.1-degree (~10 km) spatial resolution by the Climate Hazards Center.

In addition to the rainfall-based thresholds, there are several methods of confinement commonly used in the WRSI, which were also adopted here. These methods include standard growing season "windows", as well as a climatology-based planting window. The growing season windows are particularly important in areas such as Kenya, in which there is a mixture of unimodal and bimodal rainy seasons, with one of those seasons crossing the calendar year.

In east Africa, the monitoring window for the long rains season begins February 1st (dekad 4) and ends at the end of November (dekad 33). Monitoring of the short rains begins September 1st (dekad 25) and concludes at the end of February (dekad 6). The second method of confinement is an iteratively calculated, climatology-based planting window. This involves calculating the mode historical SOS_{WRSI} for each gridcell (1981–2019), and then confining SOS dates to be no earlier than three dekads before the mode SOS_{WRSI}, and no later than nine dekads (three months) after the mode. This confinement rule is then used to recalculate historical SOS dates, and the mode SOS_{WRSI}, until convergence (or 10 iterations).

Together, these two confinement methods are intended to simulate farmer behaviors; e.g., if planting usually occurs some time in February, they are unlikely to plant in July, even if an abnormal rainfall event were to technically satisfy planting conditions. Furthermore, at a certain point in a season experiencing delayed rains, a farmer will likely be forced to choose to abandon cropping for that year and search for alternative means of employment or funding (e.g., crop insurance). This also helps to limit instances in which an SOS_{WRSI} may be double counted across two overlapping growing seasons. If there is insufficient rainfall for an SOS_{WRSI} within the confined dates for a given gridcell, then the SOS for that given gridcell is No Start. The mode of the SOS between 1981–2019 in the

Table 5
Summary statistics for discrete demographic variables. SD is standard deviation. Min is minimum value. Max is maximum value.

Variable	*n*	Mean	SD	Min	25%	50%	75%	Max
Farm size (ha)	608	2.80	2.56	0.1	1.25	2.0	3.50	22.0
Household size (number of people)	609	4.28	2.00	1.0	3.00	4.0	5.00	16.0
Total Livestock Units (TLU)	608	2.48	1.97	0.0	1.21	2.1	3.37	15.4

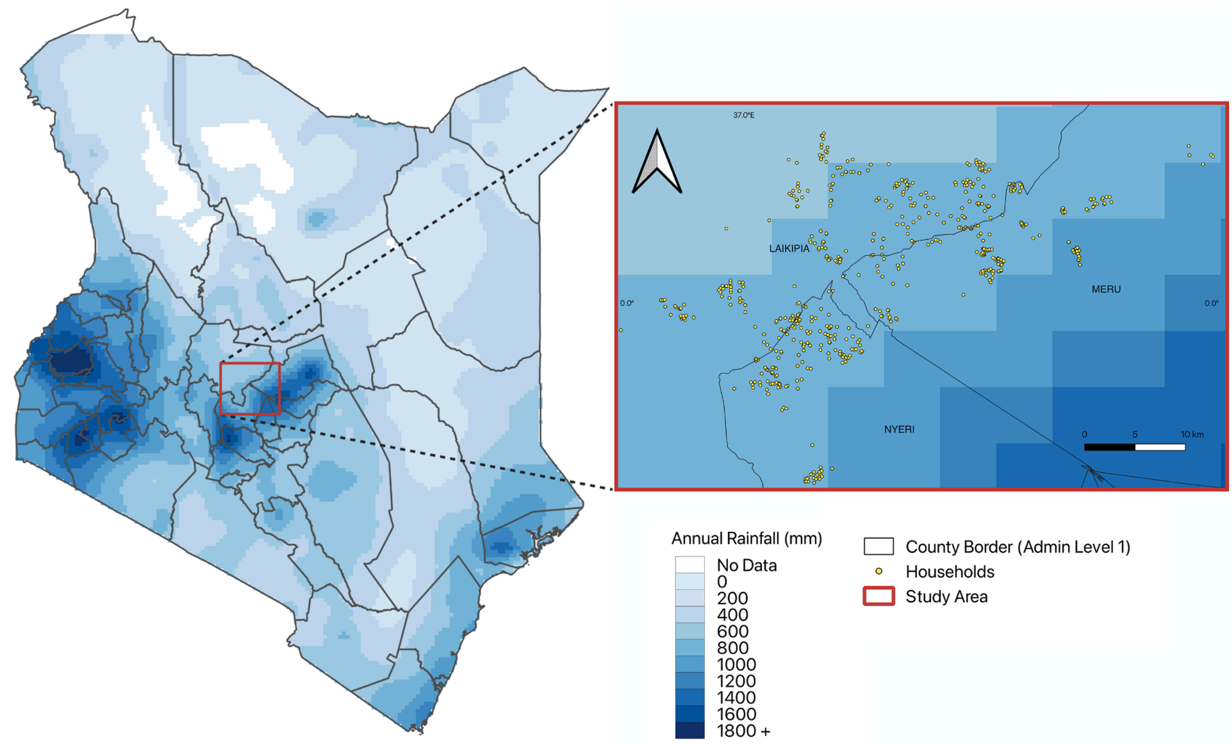

Fig. 1. Map showing study area (denoted by a red rectangle) within central Kenya. Inset map shows household locations in Meru, Nyeri, and Laikipia counties. Spatial variability in average annual rainfall is shown by the blue colorbar. Average annual rainfall was estimated using historical annual rainfall (1983–2016) from an enhanced Climate Hazards Center InfraRed rainfall with Station data (CHIRPS) product that was created by blending 710 quality controlled dekadal station observations with the publicly available CHIRPS product (Funk et al., 2015). Contours of 200 mm were delineated using the GeoCL.IM Contour Tool.

study site was between the first dekad of March and first dekad of April for the long rains and the first dekad of October for the short rains as shown C.

3.3. Variable description

We use a precise value of SOS_{WRSI} for each farmer based on the farmers' household location where their planting occurred. There is spatial variability in the gridded SOS_{WRSI} dataset in which multiple values of SOS exist for each year and season under study. Therefore, we extracted the dekad of SOS based on the farmers' household location. Both the SOS_{WRSI} and week of planting variables are in dekads, and so we subtracted the SOS_{WRSI} from the planting dekad to get an estimate of planting decisions. Negative values tell us that the farmer planted before the SOS_{WRSI} whereas positive values tell us that farmers planted after the SOS_{WRSI}. We use three EO datasets that have a theoretical relationship to farmer planting dates and the onset of the rainy season (Table 2). A farmers' timing of planting is a decision that not only relates to environmental drivers but is also governed by social and logistical constraints. Thus, we also consider household variables that are included in the random forest models as independent variables (Table 3). We use the dependent variable of planting decision and several EO datasets as independent variables. The farmer planting decision is the difference between the SOS_{WRSI} and the farmer's week of planting in dekads.

As described in Table 3, we use a subset of individual and household-level variables, selected in accordance with the literature, to capture the influence of socio-economic factors on planting date. The household-level attributes chosen are meant to provide a baseline for the demographic characteristics of the farmers rather than seeking causal drivers of planting. We report descriptive statistics for the farmer demographics in Tables 4 and 5. As shown in Table 4, the majority of farmers (74%) are part of a CWP. The most frequent levels

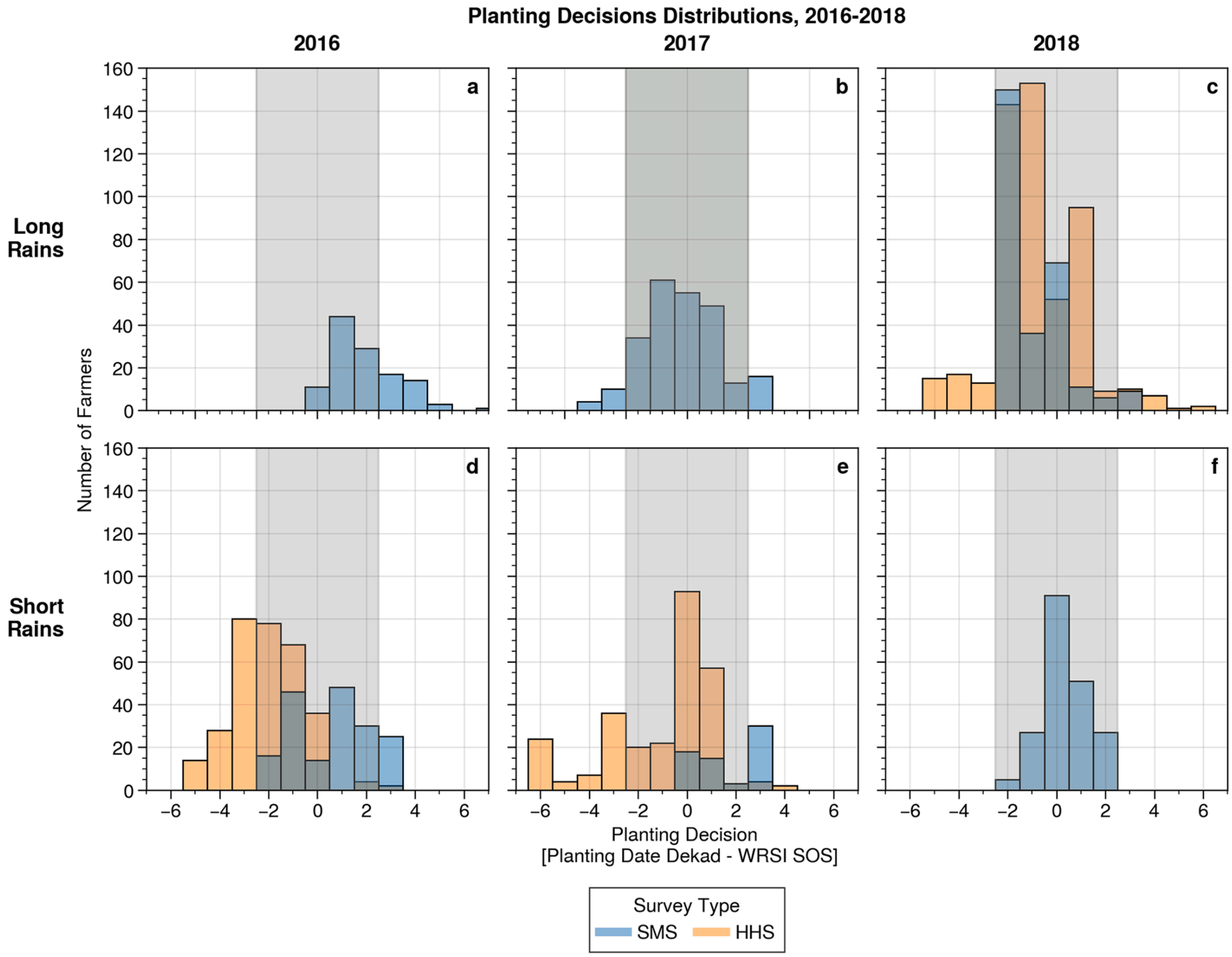

Fig. 2. Distributions of difference between planting decisions in dekads (PD) and the SOS_{WRSI} in dekads. The gray-shaded envelope between −2 and 2 indicates "on time" planting. Negative dekad values less than −2 indicate that farmers planted early and positive values greater than 2 indicated late planting. Subplots arranged by year (columns) and season (row); a-c show planting distributions for long rains; d-f show distributions for long rains. Distributions are reported for SMS surveys in blue and HHS surveys in orange. The SMS and HHS surveys overlap for three survey periods: short rains 2016, short rains 2017 and long rains 2018 (subfigures c, d, and e).

of education attainment are completed primary school (22%) and completed secondary school (22%). The households in our survey had, on average, 4.28 members, 2.48 Total Livestock Units (the methodology for calculating this variable can be found in Jahnke (1982)), and a farm size of 2.80 ha (shown in Table 5). Previous studies have further investigated the socio-demographic attributes of the farmers, such as use of weather forecasts (Guido et al., 2020; Guido et al., 2021) and use of mobile phones for agriculture (Krell et al., 2020).

3.4. Spatio-temporal patterns in planting decisions

The following description of spatio-temporal patterns in planting decisions describe Figs. 2 and 3. In Fig. 2, we characterize the distribution of planting decisions by years and seasons for the two survey types. Within those distributions, we consider whether farmers planted before the SOS_{WRSI} (which are indicated by negative values on the x-axis), after the SOS_{WRSI} (positive values) or in the same dekad as the SOS_{WRSI}. A crude estimate of "on time" planting would be when the SOS_{WRSI} dekad matches exactly with the dekad of farmers' planting. However, we acknowledge that simply defining late or early planting as 1 or more dekads away from the SOS_{WRSI} is too restrictive. Thus, in our evaluation of the temporal distributions of planting decisions, we take a conservative approach in how we evaluate late and early planting. For the purpose of visualizing how many farmers planted before or after the SOS, we consider on time planting to be farmers who planted within two dekads of the SOS_{WRSI}. In other words, farmers who planted three or more dekads (equivalent to 30 or more days) before the SOS_{WRSI} planted "early," and farmers who planted three or more dekads after the SOS_{WRSI} planted "late." Designations of "on time" planting are indicated by the vertical gray bars in Fig. 2. We use this conservative estimate of on time planting because there can be errors both in terms of the recording of planting dates and in the SOS_{WRSI} estimate. Note, these descriptions of "on time" planting are not used in the Random forest models; rather, we use the planting decision variable for modeling

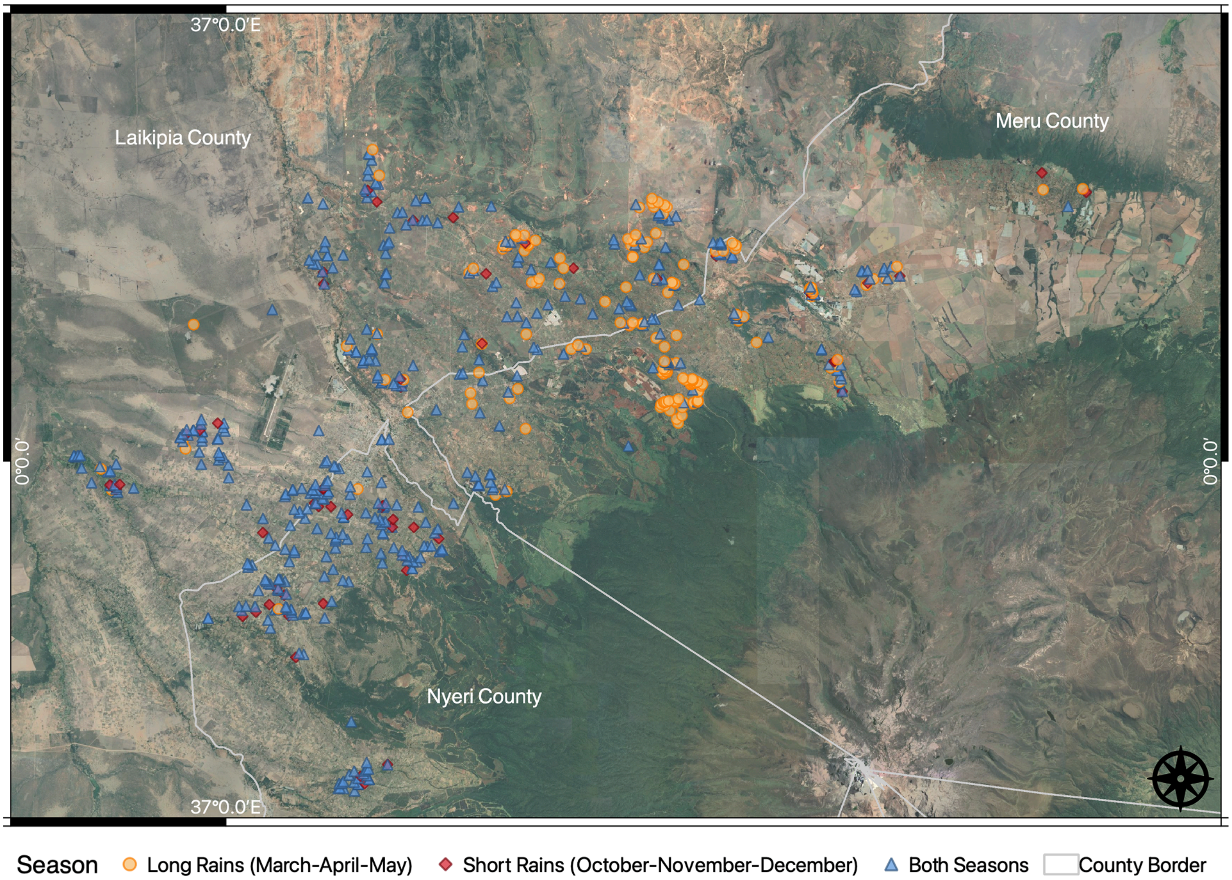

Fig. 3. Map showing spatial distribution of farmers in the study site by season planted. The transparency is set to 70% on all markers. Season planted shown as orange circle for long rains (MAM), red diamond for short rains (OND), and blue triangle if both seasons were planted.

purposes.

Because we have SMS-reported planting dates for all of the seasons, we can compare temporal patterns in planting between seasons and years. Looking at the SMS-reported planting decisions (blue histograms in Fig. 2), there is variability in planting behaviors between seasons and years. We show that planting decisions generally followed a unimodal distribution for the 2016 and 2017 long rains and 2018 short rains. In the 2016 long rains, the majority of farmers planted after the SOS_{WRSI} (indicated by positive values) whereas in the 2017 long rains season the majority of farmers planted on time. The 2018 long rains season had two peaks of planting decisions dekads, and thus shows a bimodal distribution compared to the previous two years. Both of these peaks fall within the on time designation and fewer than twenty farmers planted late. We describe deviations from on time planting behavior during the 2016/2017 rainfall seasons compared to the 2018 season in Section 6.3.

For the short rains, the distributions of planting decisions are more variable across years. Looking at just the SMS survey type, there were multiple peaks of planting in 2016 and 2017 whereas in 2018 there was one peak. In 2018, the majority of farmers planting in the same dekad as the SOS_{WRSI}, and the planting decisions are normally distributed around that behavior. In the 2016 short rains, most of the farmers planted on time. In 2017, farmers planted on time or late but within three dekads of the SOS_{WRSI}. In 2018, all farmers planted on time. Because the SMS-reported planting decisions are the focus in the rest of the paper, we omit a discussion on the differences between SMS and HHS-reported planting decisions which can be found in Appendix D.

We show the spatial distribution of farmers' planting patterns by season in Fig. 3. Across the landscape, farmers in all counties plant during both rainy seasons, as indicated by the blue triangle symbol. On the eastern side of the study site, such as on both sides of the Meru and Laikipia county boundary line, we see that farmers slightly favor planting only once per year, during the long rains. Farmers who only plant during the short rains (OND) are more likely to be in Laikipia or Nyeri counties compared to Meru county.

4. Methods

We use random forest models to evaluate how well EO products that are typically used for agricultural modelling can be used to predict whether farmers plant early, on time, or late relative to a rainfall-based estimate for the start of season in dekads.

Prior to model fitting, we convert each of our environmental predictor variables into anomalies (value in a given month differenced from a mean calculated from prior years). The use of anomalies as predictors is standard practice when working with EO variables as

Table 6
Summary statistics of planting dates and SOS_{WRSI}. Data includes two survey types (SMS and HHS), three years (2016–2018), and two seasons (MAM and OND). *n* represents the number of farmers providing their planting date for that season, year, and survey type. We report the mean, median, and standard deviation of the planting date in dekads. Planting range (inclusive) represents the earliest and latest dekad of planting reported. SOS_{WRSI} mean is the average SOS in dekads experienced among the farmers since there is spatial variability in the SOS_{WRSI}. Because certain seasons had more than one start of season due to the spatial heterogeneity of rainfall, we also note the number of unique number of SOS_{WRSI} in the study site per season.

Survey Type	Year	Season	*n*	PD Mean	PD Median	PD Mode	PD SD	PD Range	PD Unique	SOS_{WRSI} Mean	SOS_{WRSI} Unique
SMS	2016	MAM	119	10.94	11.0	10	1.34	[9, 16]	7	9.00	1
		OND	179	30.69	31.0	29	1.48	[28, 33]	6	30.10	2
	2017	MAM	242	9.72	10.0	10	1.53	[7, 14]	8	9.93	2
		OND	66	29.68	29.5	31	1.30	[28, 31]	4	28.00	1
	2018	MAM	281	7.98	7.0	7	1.28	[7, 12]	6	9.00	1
		OND	201	28.69	28.0	28	0.79	[28, 30]	3	28.35	3
HHS	2016	OND	310	28.06	28.0	28	1.40	[25, 33]	9	30.10	2
	2017	OND	272	26.97	28.0	28	2.23	[22, 32]	11	28.00	1
	2018	MAM	517	8.16	8.0	8	1.74	[4, 15]	12	9.00	1

Table 7
Description of model specifications.

Model Name	Model Description
[1] survey	Uses only survey variables
[2] precip	Uses survey variables and the CHIRPS rainfall product
[3] NDVI	Uses survey variables and NDVI
[4] ET0	Uses survey variables and evaporative demand
[5] EO	Uses only EO variables (precip, NDVI, and ET0)
[6] all	Uses all EO variables and all survey variables (combines models [1] and [5])

they place the key variable in a context relative to spatial and seasonal averages. In this paper, we explore if the reference period used to define an anomaly impacts model performance. Specifically, we define anomalies over 1,3,5, and 10 year rolling averages and compare predictive performance of these anomalies across models. For example, if the performance of an EO product based on a one year anomaly notably outperforms a model based on a 10 year anomaly, this provides useful information for future planting models and also suggests that farmers are more likely to respond to more recent, rather than historical, weather trends when making decisions about planting dates. Formally, the anomalies are defined by:

$$\widetilde{EO}_{(m,y)} = EO_{(m,y)} - \frac{1}{N}\sum_{y-N}^{y-1} EO_{(m,y)} \tag{1}$$

where EO is an earth observation product listed in Table 2, m indexes calendar months, y indexes years, and N refers to the 1,3,5 and 10 year periods.

We compare the forecast accuracy of these different anomaly definitions in several different model specifications. Each model specification includes various combinations of survey and EO variables (Table 7). Specifically we compare in and out-of-sample forecast accuracy of a model that uses just the survey variables (model [1]), models that use survey variables along with one of the EO products (models [2,3,4]), a model that uses only EO variables (model [5]), and a model that uses all survey and EO variables (model [6]). We cleaned and prepared the planting data for analysis using Python programming language (Python Software Foundation et al., 2020). The random forest models were completed using the randomForest and rpart packages in R (Liaw and Wiener, 2002; Therneau et al., 2019).

We fit separate models for the HHS and SMS survey types but focus our primary results on the SMS surveys since the planting dates in these surveys are not based on recall and thus tend to be more accurate. We also fit separate models for the county of Lakipia and for the combined counties of Meru and Nyeri because this provides roughly equal sample size between the two partitions. To test out of sample accuracy we use the model fit for Laikipia to predict Meru/Nyeri, and vice versa. Our primary results focus on models that are pooled across years. However, because EO variables tend to have more temporal variation than household variables–and thus might have a predictive advantage over survey variables–we also report supplementary results for separate models run on each year.

5. Results

The results of the random forest models are shown in three figures. Fig. 4 compares overall performance across models [1–6] and anomaly periods (1, 3, 5, and 10 year). Fig. 5 compares the performance of individual variables and Fig. 6 examines the out-sample-predictive accuracy. Corresponding figures displaying the results for each individual year are shown in Appendix F.

We begin with Fig. 4 to explore the degree to which survey and EO variables can be used to predict the planting date anomalies of small-scale maize producers. Fig. 4 shows in-sample Mean Squared Error (MSE) values) for models [1–6] with the four different reference periods (1, 3, 5, and 10 year) used to define EO anomalies. The results are separated by models fit for each region (Laikipia

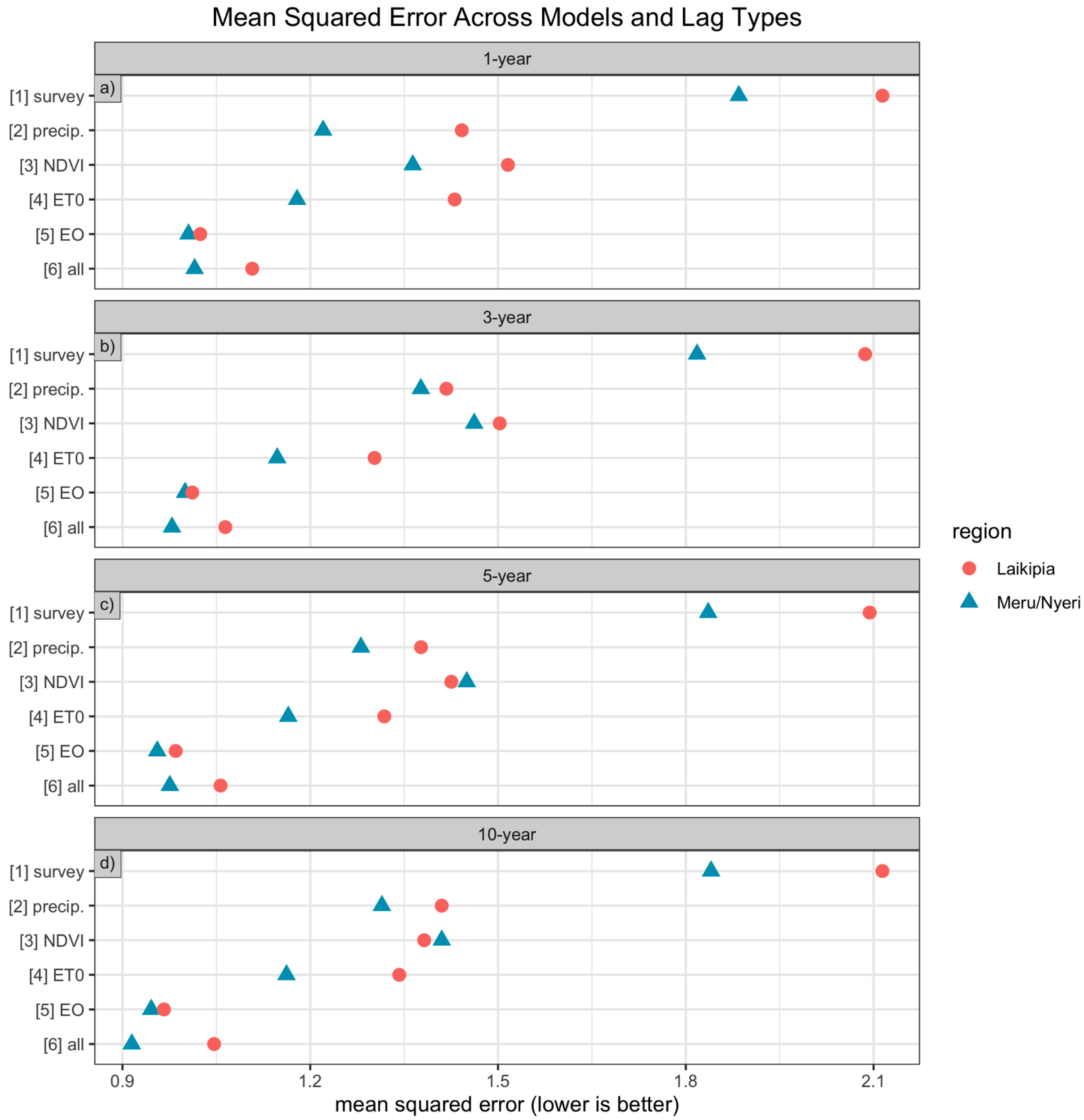

Fig. 4. Random forest model results showing Mean Squared Error (MSE) across models (rows within subplots, 1:6) and lag types (rows of subplots a: d). Results are subsetted by region where Laikipia county is demonstrated by the circle symbol; Meru and Nyeri counties are demonstrated by triangles.

and Meru/Nyeri).

The best performing models across all lag types are the models with solely EO variables (Fig. 5, 5-EO) and the models with all of the variables (Fig. 5, 6-all). Models [5] (EO only) and [6] (all variables) perform similarly across lag types and regions: the model with only EO variables [5] had MSE values between 0.92 and 1.02 across lag types and regions while the [6] model with all variables had MSE values between 0.97 and 1.11. Overall, the models with EO datasets (5-EO) and the models with all of the variables (6-all) exhibit the strongest predictive relationship with planting behavior as demonstrated by the low MSE.[1]

We find that models with only EO variables [5] perform better than models with only survey variables ([1] survey) or any individual EO model ([2] precip, [3] NDVI, and [4] ET0). However, when looking closely at individual EO datasets, we find that not all EO datasets perform similarly. We find that the rainfall and ET0 models have stronger influences on planting decisions compared to NDVI. Contributions of individual variables to model fit are presented in Fig. 5.

[1] While the focus of our paper is on out of sample predictions we also include the corresponding plots with R2 values-squared in E). The values shown in E.9 correspond with those here: models [5] and [6] have the highest explained variance with values ranging from 0.6 to 0.7.

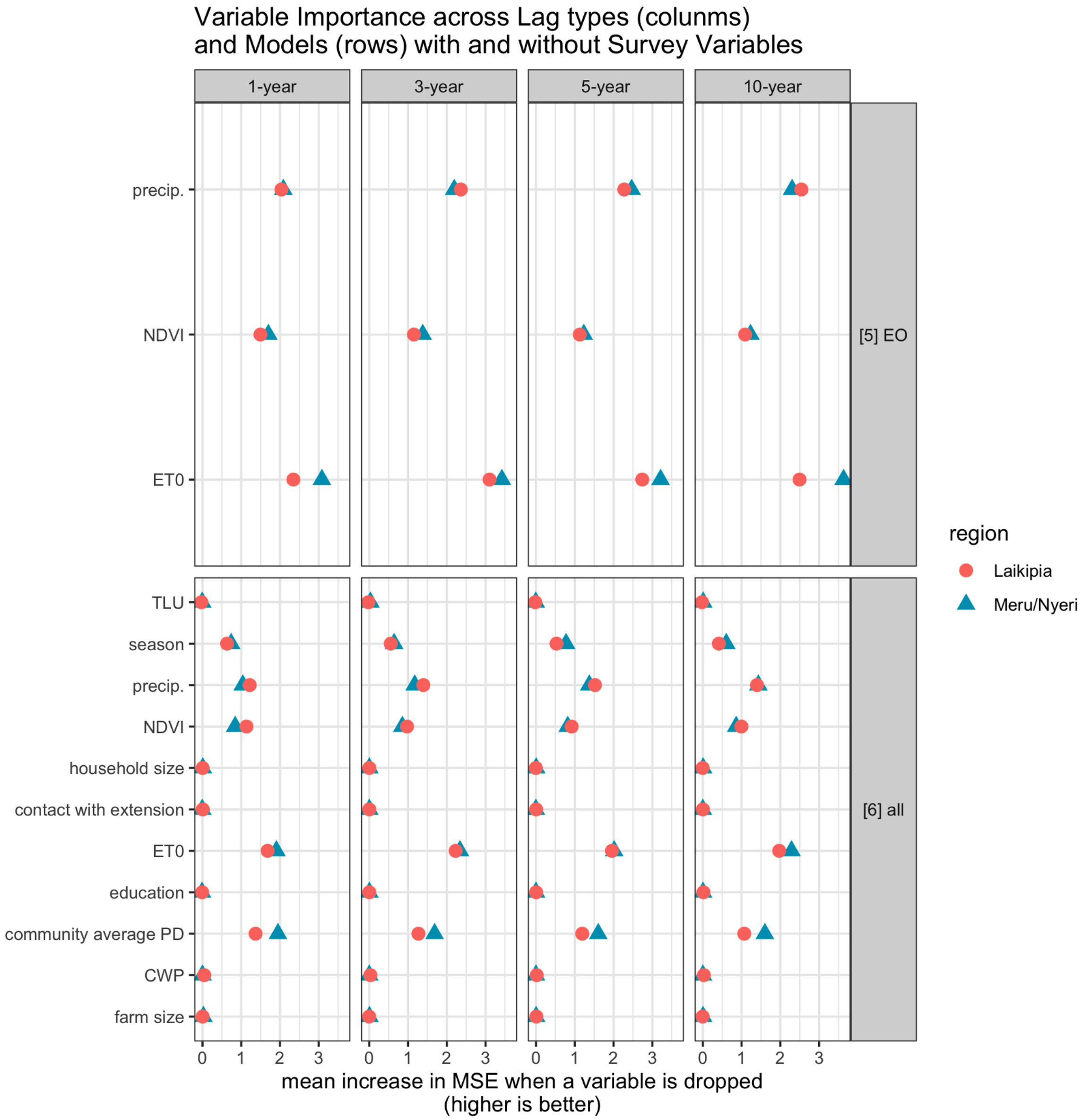

Fig. 5. Random forest model results showing variable importance. Higher values in mean increase in MSE when a variable is dropped means higher importance of a given predictor variable. Lag types are presented as columns (a:h) and variables within models are presented as rows. Results are subsetted by region where Laikipia county is circle symbol; Meru and Nyeri counties are triangles.

In Fig. 5, we disaggregate the results to analyze variable importance in the best performing models: [5] and [6]. We use the standard measure variable importance for random forests: the mean increase (across all trees) in MSE when a given variable is excluded from a model. By this methodology, rainfall and ET0 tend to contribute most to in-sample predictive accuracy regardless of model or lag type. Across lag types (1,3,5, and 10-year anomaly reference periods), rainfall exhibits a marginal increase in importance with longer than 1-year lag type. More notable is an increase in ET0 importance, particularly at Laikipia, with a longer than 1-year lag type. This suggests that evaporative demand anomalies defined over longer reference periods may be stronger predictors of planting dates compared to ET0 differences from the previous year.

The community average planting dekad is the highest ranked variable of the survey variables. This is unsurprising as this variable is a spatial semi-autoregressive variable in which the dekdad of planting date is included in the dependent variable. However, ET0 is ranked equal to higher than community average planting dekad in every model and for both regions. This indicates that ET0 has the potential to be a robust predictor of planting decisions. These results also hold when comparing variable importance on separate models run for each year as shown in Fig. F.11.

Finally, we explore the out-of-sample prediction errors in order to provide some indication of how the results might generalize to the other areas. Fig. 6 shows the out-of-sample MSE for the two highest performing models. For example, in the “Meru/Nyeri” row, a

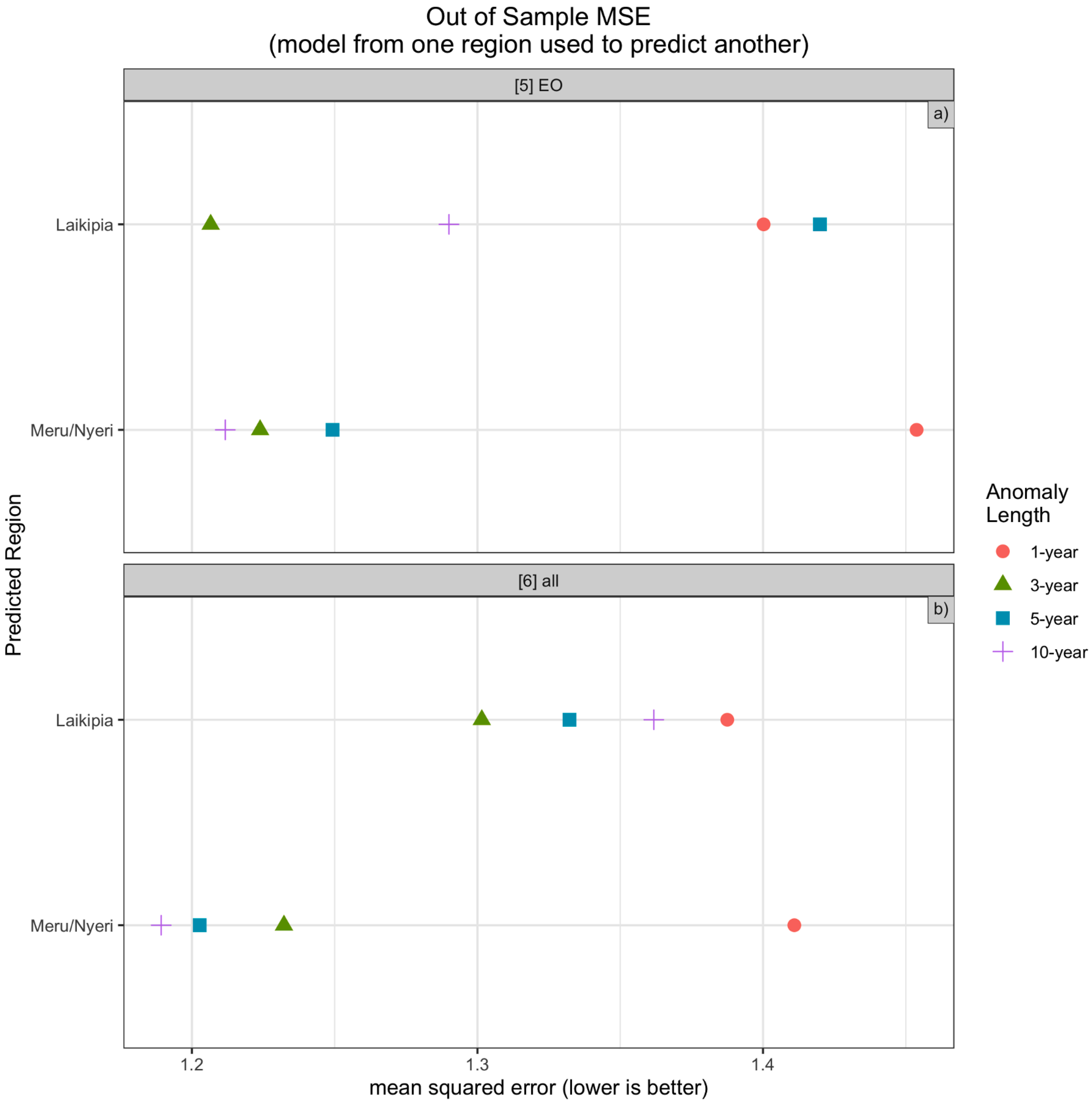

Fig. 6. Out of Sample MSE results. Plot showing models 5-EO (top panel) and 6-all (bottom panel) Out of Sample MSE by region (rows) for the SMS survey type.

model was fit to the Laikipia data and then used to predict Meru and Nyeri households. Overall, the out-of-sample accuracy does about as well as in-sample accuracy for the two top performing models.

We do find some difference in the out-of-sample performance across lag types. The largest difference is found in 10-year lag. The out-of-sample MSE for Laikipia (i.e. the model fit to Meru and Nyeri and then used to predict Laikipia) is higher than the out-of-sample MSE for Meru/Nyeri for the longer lag types (5- and 10-year), whereas the reverse is true for lag type of 1-year. Overall this suggests that EO variables may be able to serve as predictors for planting decisions, but the time period used to define the reference period may be more idiosyncratic and vary by village and even by farmer.

6. Discussion

6.1. Planting decisions relate best to EO data, specifically rainfall and ET0

We find that rainfall and evaporative demand relate best to planting decisions out of the EO datasets explored. Intuitively, rainfall and ET0 should predict planting decisions relatively well since these are environmental variables that farmers recognize and respond to. While farmers likely do not have instruments on their farm that measure rainfall or ET0, they sense the temperature and experience changes in water availability around the time of planting. Rainfall alone does a fine job of characterizing planting decisions as indicated by the majority of farmer-reported planting dates falling within a +/- 2 dekad window of the rainfall-based SOS_{WRSI} estimates for planting date. Previous work has characterized the relationship between rainfall and planting behavior in rainfed settings (Marteau et al., 2011; Sultan et al., 2005) and concludes that rainfall, specifically the onset of the rainy season, is the key variable for producing optimal yields in rainfed agricultural settings (Cammarano et al., 2012; Mugalavai et al., 2008; Tang et al., 2018). Thus, the relationship between rainfall and planting decision is not surprising, and further supports the existing practice of using rainfall in modeling studies and famine early warning applications to provide large-scale trends in planting date decision-making.

We find that ET0 performs very well as an EO product that can be used to predict planting decisions. ET0 is related to farmer planting decisions such that with increased solar radiation, farmers likely recognize if it is hotter and drier, and thus the need to delay their planting. While the importance of ET0 increases with the length of the lag used to define the anomaly, ET0 even at the 1-year lag has a mean increase in MSE above 2 when ET0 is dropped (Fig. 5). Finding a strong positive association between ET0 and planting decision is a contribution to the literature on the environmental drivers of planting dates. ET0 is a function of several meteorological drivers including incoming shortwave radiation, net longwave radiation, temperature, specific humidity and wind speed. For our study site in Kenya, incoming solar radiation influences ET0 more so than other drivers (Hobbins et al., 2018; Harrison, 2018). Our results show that rainfall is not only related to planting dates but ET0 provides additional information that could be better utilized and is discussed in Section 6.4.

The relationship between planting and NDVI is not very clear from our model results. This could be due to a number of factors related to the heterogeneity in land cover and crop phenology at the field-scale. NDVI at a ~8 km pixel resolution may not be sensitive to planting date and vegetation green up for a number of reasons. Spatially, the landscape is heterogeneous and consists of natural evergreen vegetation, natural grasses, and shrubs whose greenness is affected by rainfall and rainfed agriculture. However, even if the landscape was uniformly cropped, the greenness signal may not be significant after only 1 month after planting.

Experimenting with different NDVI methodologies could be informative. This study examined monthly NDVI, to see if the departure from average during the rainfall-derived potential onset month could indicate a clear signal for delayed or early vegetation growth. Focusing on sub-monthly NDVI and for a longer window of time could provide more insights to rainfall-vegetation growth lags and NDVI performance for estimation of planting dates in this region. Investigations could estimate start of season dates from the upward curve of function-fitted NDVI time series, which is a technique used in many studies (e.g. Heumann et al. (2007); Zhang et al. (2003); Eklundh and Jönsson (2016)). A useful software package for this is TIMESAT (Jönsson and Eklundh, 2004), which can assist with the multi-step process of time series filtering, function-fitting, and estimation of phenology characteristics from remotely-sensed data. Future studies could further investigate these approaches such as use of TIMESAT.

6.2. Relationship between planting dates and household characteristics

Overall we find poor correspondence between planting dates and household variables in our model. EO datasets are, overall, stronger indicators of planting decisions; however, this is not to say that socio-economic variables are irrelevant. Rather, increased research on the socio-demographic determinants of planting is needed. The inability of our models to predict planting date using household survey data could be a reflection of i) the inherent heterogeneity of households in the study site and ii) the lack of true insight into household-level determinants of planting. In this region of Kenya, households exhibit a large degree of heterogeneity in income, education, family and farm size, irrigation access, and participation in agricultural cooperatives and water resource governance groups (Giroux et al., 2020; Gower et al., 2016; Krell et al., 2020).

We show that environmental markers better explain planting decisions despite the variability of such markers. Because we do not make large gains in model performance by including household survey variables, further research–likely qualitative in nature–is needed to investigate what household factors strongly influence planting decisions. While we capture some of these attributes in our household surveys, further work using Randomized Controlled Trials (RCTs), focus groups, and/or other qualitative means of data collection could better reveal the types of variables that should be included in these models. Perhaps the variables selected in our study did not function well enough as proxies and other variables that were not observed would have performed better. In general, a deficit in the literature exists about the determinants of planting. The impacts of socio-economic factors, community behaviors, and logistical and time constraints in rural east African settings on planting decisions ought to be further investigated at the household-level. Future research is needed to understand why household variables correlate so weakly to population-wide planting decisions in this context.

6.3. Temporal variability in planting decisions

Our real-time SMS-collected data on the week of planting shows that farmers planted within 20 days of the SOS_{WRSI} for the majority of the seasons (Fig. 2). The seasons in which farmers fell outside of the bounds of 20 days were the 2016 long rains and the 2016/17 short rains, whereas all farmers planted within the 20 days of the SOS in the 2018 long and short rains. We proceed with describing seasons in which farmers' planting corresponded with the SOS as well as those which deviated from on-time planting as well as possible explanations for those trends.

Many regions in Kenya experienced drought between 2016 and 2017. Specifically, the short rains in 2016 and the long rains in 2017 were particularly dry in our field site in central Kenya (Funk et al., 2017; Funk et al., 2018; Uhe et al., 2016). It is noteworthy that in two of the seasons which experienced drought conditions, many of the farmers still planted on-time relative to the SOS_{WRSI} for that year, which is possibly in response to waiting for the rains to come. It is possible that farmers were eager to plant and in their anticipation to plant had prepared their fields to align planting with the SOS. Some of the farmers planted late in the long rains of 2016, which could be related to the volume of rainfall during that season as well as an adaptation strategy.

Farmers' response to weather conditions by planting late is an adaptation strategy. Both planting date and cultivar choice matters in the context of a growing season as they can shift the flowering time of the crop to part of the season which may have heat and temperature conditions that results in improved yields. However, planting late can leave cereal grains such as maize susceptible to mycotoxins contamination before harvesting (Blandino et al., 2017). Previous work investigating adaptation options among sorghum farmers in west Africa showed that late sowing provided little benefit (Guan et al., 2017). Despite not knowing the outcome of their production, we show that farmers tended to plant late during the long rains of 2016 and the short rains of 2017. Importantly, we found that very few farmers planted during the 2017 short rains season. Many regions experienced rainfall deficits that led to crop water stress during this season (FEWS NET, 2017) and our real-time SMS data indicated that few farmers planted (Appendix D).

There are multiple causes for early planting. In the long and short rains of 2018 almost all of the farmers planted within +/-2 dekads of the SOS_{WRSI}. The 2018 long rains was unique as the wettest long rains season in over 100 years (Kilavi et al., 2018). The 2018 long rains season was abnormal in that the rains caused flooding, land subsidence (Rateb and Hermas, 2020), and displacement (MacLeod et al., 2020). Apart from abnormally high levels of rainfall, farmers may plant early due to a mismatch in timing or availability with weather forecasts. Patt et al. (2005) assert that farmers may only hear of forecasts after they have made planting decisions. Additionally, false onsets can drive early planting behavior. False onsets are periods of heavy rainfall storms followed by dry spells or dry periods, which can make it seem that the start of season has occurred when it has not (Dunning et al., 2016). False onsets pose a risk to agriculture because significant crop loss can occur if timing of planting is misaligned with rainfall and thus germination cannot occur or seedlings cannot survive during the subsequent dry period (Dunning et al., 2016; Marteau et al., 20090. Lastly, some farmers may decide to plant early in order to double crop within that season (Borchers et al., 2014).

6.4. Significance of findings for modeling and Famine Early Warning Systems applications

Planting dates are important variables for crop modeling studies as they are instrumental in determining the climatic conditions experienced for crop growth. Similarly, the onset of the rainy season is also used as a unique variable for drought early warning systems applications (Shukla et al., 2021). In addition, delays in the rainy season onset have a strong relationship to grain prices which can be used to enable famine early warning efforts (Davenport et al., 2021). Our study combines both planting date and the season onset in order to evaluate how well farmers' decisions align with EO and household variables. The finding that both rainfall and evaporative demand have the strongest predictive relationship to planting decisions is instructive for Famine Early Warning Systems applications which can use these EO products for triggering relief intervention. Additionally, changing the reference period used to define the EO anomaly only leads to small changes in predictive accuracy in the random forest models; therefore showing that even having the previous year's EO data is instructive for estimating farmer planting dates. Guido et al. (2020) showed previously that many Kenyan agriculturalists will form their expectation of upcoming seasonal rainfall on their observations of past, specifically recent, rainfall seasons. Therefore EO products that are recent, even from the past year, are useful for predicting farmer decision-making such as planting dates and increasing the time period used to define the anomaly does not add increasing levels of predictive power.

6.5. Study limitations

There are limitations in combining human decision-making information and EO datasets as well as sources of error within each dataset. First, many of the farmers surveyed represent a fairly homogeneous area in terms of climate. The results would have been more robust if similar surveys existed that capture real-time dynamics of planting in other semi-arid regions in Africa. Additionally, for farmers who experience higher annual rainfall totals, such as those in Meru county in the northeast area of the study site, the rainfall EO data may not be as good of a predictor for planting decisions because there is generally sufficient rainfall for planting.

Several definitions of the Start of Season exist, and there is not one single best metric for determining rainfall onset in east Africa. We used the SOS_{WRSI} because it is common in modeling studies and for famine early warning systems. However, there are some well known issues with the AGRHYMET rainfall-threshold methodology that should be acknowledged. The current implementation of a univariable rainfall-based trigger for onset of season is unable to account for sudden changes in conditions during the initial delicate seedling stage (e.g. sharp increases in temperature and wind speed, or humidity, which could result in significant soil respiration offsetting the first rain of the season). Conversely, in semi-arid regions where rainfall is low, but evapotranspiration is relatively lower, the rainfall-based onset trigger may never be satisfied, though crops are being produced, thus resulting in an underestimation of crop production by the model. The latter example is a commonly observed issue when using the WRSI model in marginal cropping zones of Somalia and Ethiopia, particularly during the short rains season. Furthermore, the rainfall thresholds currently used are somewhat arbitrary, and not definitively agreed upon (other common thresholds include two consecutive dekads of at 20 mm). In either case, it is reasonable to imagine different crops require different soil moisture levels for germination, and thus a crop specific threshold may provide superior SOS accounting. Because our definition of 'on time' planting decisions was expanded to planting within two dekads (20 days) of the SOS_{WRSI}, we have adjusted for slight variations in the start of season which may be demonstrated in other metrics of SOS.

Random forest models provide many advantages when demonstrating relationships between a set of outcome and independent variables; there are however also a few limitations. First, random forest models will overestimate continuous variables relative to discrete variables because split points are fixed for discrete variables whereas continuous variables are not. Because our continuous variables do not yield similar results to the discrete variables (e.g. household size and education level of respondent), this limitation is less concerning. Second, a reason why EO variables outperform household variables could be because EO variables change from year-to-year whereas household variables remain constant.

7. Conclusion

When to plant is one of the most important decisions a farmer makes at the start of the growing season. Often researchers and humanitarian actors are limited by empirical data on planting dates in rural settings and have to resort to index-based or satellite estimates of planting which may not be precise or adequately represent within-community variability. Using a real-time estimator of planting dates, SMS surveys, we investigated the relationship between several EO datasets and household variables for a study population of ca. 650 farmers in central Kenya. We find that planting dates vary across seasons and years and do not always co-vary with the SOS_{WRSI}. Outputs from random forest models demonstrated which combinations of socio-economic variables and environmental datasets best explain planting decisions. We find that models with only EO variables perform as well as models with both socio-economic variables and EO variables. Thus, with only the EO data (NDVI, evaporative demand, and rainfall) we can predict planting behavior fairly well. We also consider which EO variables best predict planting dates and find that evaporative demand and rainfall improve predictions more so than NDVI.

Overall, we do not find household survey variables to predict planting decisions at the population scale; however, that does not render socio-economic variables as irrelevant. Rather, while Earth Observation variables provide a near-real time snapshot of environmental conditions at relatively fine spatial and temporal scales, household level variables collected via household surveys often have a lag in when they become available and thus may be more limited in pairing with time-sensitive environmental decisions such as planting. Researchers exploring the human dimensions of planting time have primarily used quantitative analyses of household survey data to describe associations between planting date and various socioeconomic variables. Unfortunately, these existing studies are limited in their ability to explain the complex mechanisms underpinning these relationships, how socioeconomic factors interact with each other (and biophysical factors) to produce variable planting dates, and the extent to which their influence on planting time varies across scales and between places and agrarian systems. More research in this area is certainly needed. Our study improves the understanding of biophysical and household-level determinants of planting decision-making from empirical observations, which are difficult and costly to collect. The decision of when to plant is an important component of crop modeling studies and this study provides empirical evidence of the relationship between environmental factors and farm-level decisions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank the study participants for their time and engagement in the study for multiple years as well as field assistants and enumerators from the Mpala Research Centre who helped execute the work. The research was supported by the National Science Foundation Awards SES-1360421 and SES-1360463 and the U.C. Santa Barbara Earth Research Institute Graduate Summer Fellowship. The Institutional Review Board at U.C. Santa Barbara approved ethical clearance for the research. We also thank Juliet Way-Henthorne for providing professional editing assistance.

Appendix A. SMS surveys

We began the SMS program in Kenya in April 2015. We recruited farmers from CWPs that we had worked with previously and held group trainings to inform them about the SMS program. Kenyan research assistants ran these group trainings and explained how the SMS program would collect information about their farming practices for research purposes. From these trainings approximately 200 farmers decided to enroll in the program. We compensated farmers for their voluntary participation. Farmers received talk time in exchange for them answering the SMS surveys (approximately 0.20 US Dollars per week at that time), which was delivered directly to their cell phones.

We expanded farmers in the program using CWPs as an outreach method between 2015 and 2016. With the permission of the CWP managers, we visited CWP and other community-based meetings and invited all attending farmers to participate in the program. Research staff enrolled farmers from both new CWPs and non-CWPs areas to cover the region depicted in Fig. 1. In July 2017 when active recruitment for participants ended, the enrollment was 645 farmers.

To collect data via SMS, we used TextIt (https://textit.in), which is a low-cost program that allows users to design and disseminate questionnaires. Our survey was composed of a set of 4–8 questions depending on the season and farming circumstances that required no more than three minutes to answer the entire survey. We sent surveys each week on Saturday during the planting, growing, harvest, and interseason periods. In the survey questions sent during the planting season we asked farmers: “Have you planted maize in the last 7 days?” The surveys were constructed using skip logic based on the farmers’ responses; e.g. if a farmer responded that she finished harvesting all of her maize, the questionnaire automatically switched to the interseason questionnaire. Additional information on the design of the SMS survey program, which was concomitantly done in Zambia, can be found in Giroux et al. (2019).

Appendix B. Household surveys (HHS)

We completed two household surveys in 2017 and 2018 to retroactively collect information on farmers’ planting dates. We aimed to include the farmers from the SMS program in our household survey as much as possible. All respondents agreed to participate in the study through informed and voluntary consent and were above the age of 18. Respondents were not compensated for their participation. The surveys included a section on maize planting in which the farmers provided the month and week of planting maize per season. The maize planting section occurred in the first 20 min of the survey and was followed by sections on demographics, perceptions of climate and weather, use of mobile phones and weather forecasts and migration. The household surveys were conducted in either Kiswahili, Kikuyu or Kimeru depending on the respondent's vernacular language.

Between March and April 2017, we conducted a household survey using Qualtrics software (Qualtrics and Qualtricsualtrics and Qualtrics, 2019). We first sampled a representative number of households from 25 CWPs from 5 Water Resources User Associations (WRUAs): Nanyuki, Likii, Ngusishi, Timau, and Ngare Nything. Further description of CWPs and WRUAs can be found in Hannah et al. (2021). For CWPs with fewer than 15 households in its membership, we aimed to survey all households. For CWPs with more than 15 households, we randomly selected and interviewed approximately 30% of the households using a list of households provided by the CWP's chairperson. To randomly select households, we had a local contact from each community create a list of the households in the CWP and we set up a meeting with a subset of those farmers to ask if they would like to participate in the survey. We surveyed farmers who participated in the SMS program as much as possible. We also included households that were in the same geographic area as the CWP being sampled that day but were not participating in the CWP. Therefore we expanded the survey to an additional 32 CWPs. We surveyed 504 farmers, but only used responses from 435 farmers for whom we had both planting date information and household coordinates. We removed respondents who were flagged as giving incomplete answers. Further details on the sampling strategy can be found in (Waldman et al., 2019).

Between June and July 2018, we conducted a follow-up household survey. We revisited as many of the CWP and non-CWP households as possible from the 2017 survey. For households that we could not get in touch with or were unwilling to participate in our study, we randomly sampled neighboring households within the same community. We interviewed 605 respondents; however 516 respondents were used in the study because not all respondents reported planting maize. Additional details of our data collection and other variables included in the household survey appear elsewhere (Guido et al., 2020; Krell et al., 2020).

Appendix C

Fig. C.7 shows the SOS_{WRSI} mode between 1981–2019 for the long and short rains.

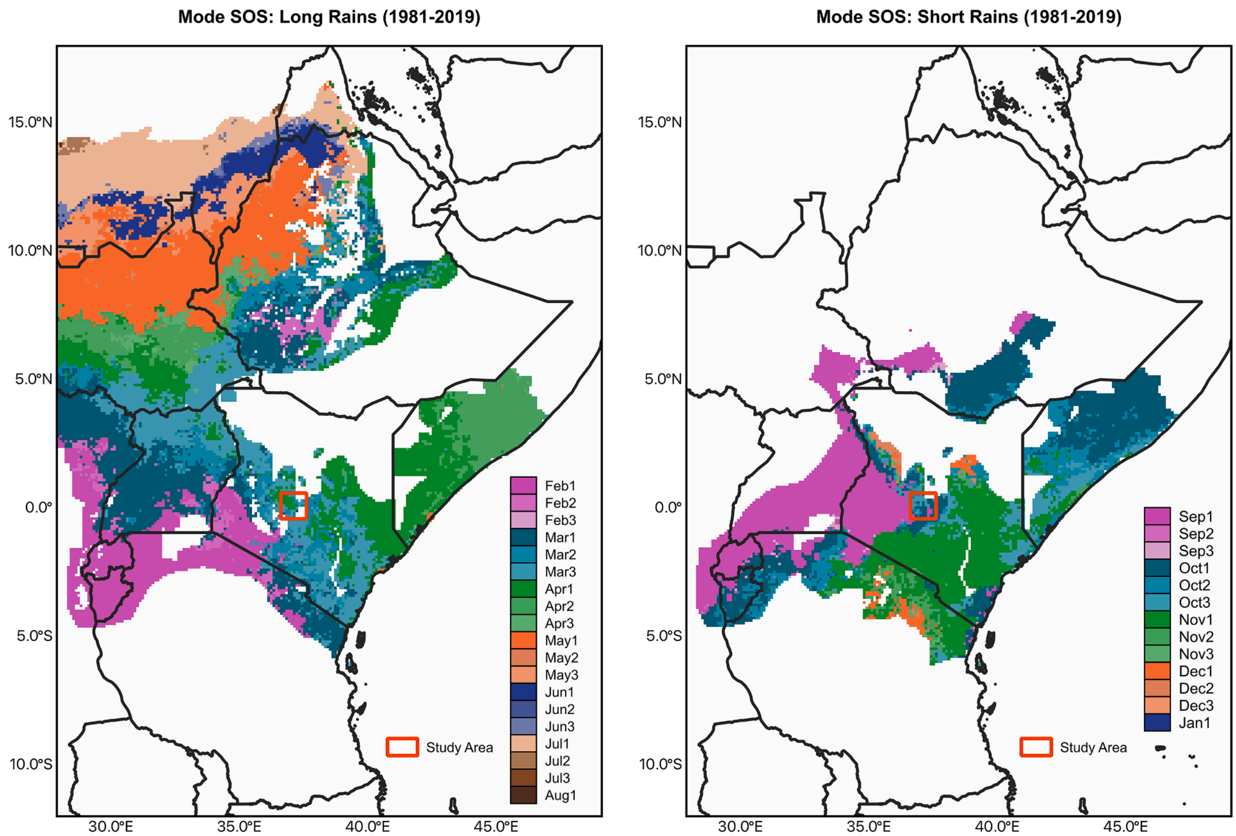

Fig. C.7. SOS_{WRSI} mode between 1981–2019 in the region. Long Rains is shown in the figure on the left and the Short Rains is shown in the figure on the right.

Appendix D. Differences between SMS and HHS survey types

We can compare differences in planting decisions between the two survey types for three seasons: short rains 2016 and 2017, and long rains 2018. We are interested in the farmers who responded to both survey types—SMS and HHS—and have plotted the planting distributions in Fig. D.8. Visually, we can identify differences in planting decisions between the two survey types for the 2016 OND and 2018 MAM seasons. We will omit the 2017 OND season from this discussion due to the low number of samples. The HHS data show earlier planting and a longer range of dates than the SMS data. For the long rains 2018 season, the planting distribution matches fairly well between the two survey types; however there is a greater range in values for the household survey (HHS) planting decisions compared to the SMS data. For this season, both survey types showed that the majority of farmers planted "on time"; however for the HHS-reported data there is a wider range of planting dates recorded. For the short rains in 2016, we again see earlier planting on average from the HHS responses. The HHS-reported planting decisions are approximately 3 to 6 dekads earlier than the SMS-reported dates.

Because of the differences between the SMS and HHS-reported planting dates, we only use the SMS-version of the farmer planting dates in our random forest models. These differences could be due to the recall-based nature of the HHS in which these planting dates were asked about between 6–12 months after planting. Uncovering the reasons why the two survey types result in different planting decisions patterns is beyond the scope of this study, but warrant further investigation.

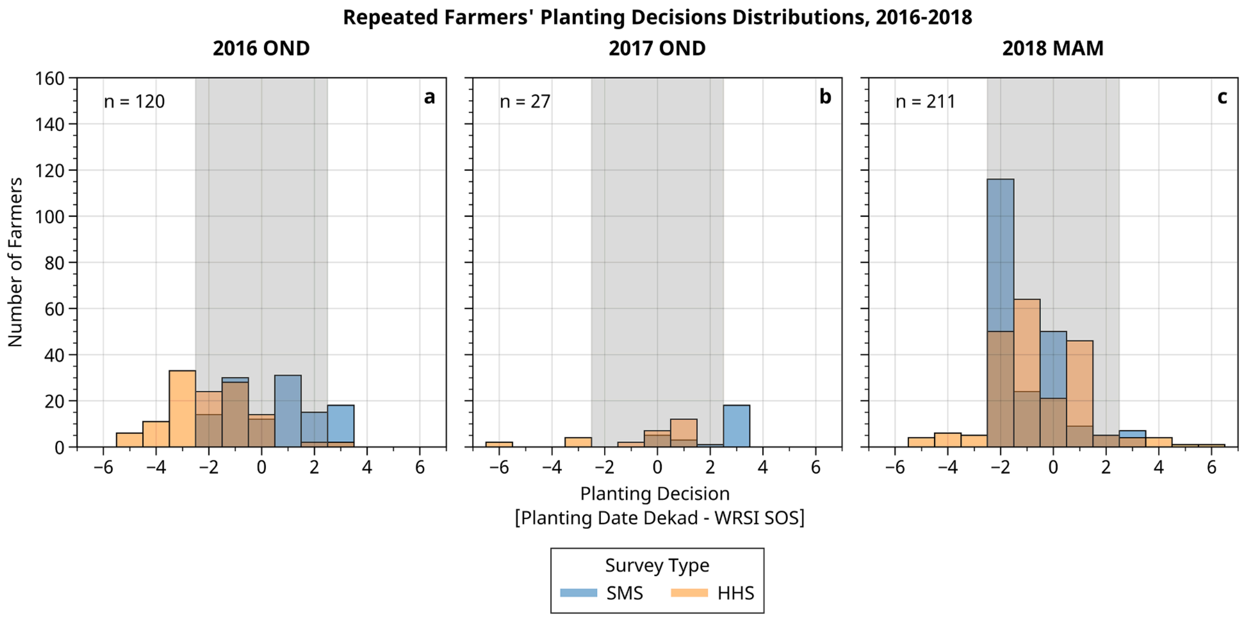

Fig. D.8. Planting decisions distributions for farmers who responded to both survey types (SMS and HHS).

Appendix E

Fig. E.9 shows the R-squared values for the random forest models.

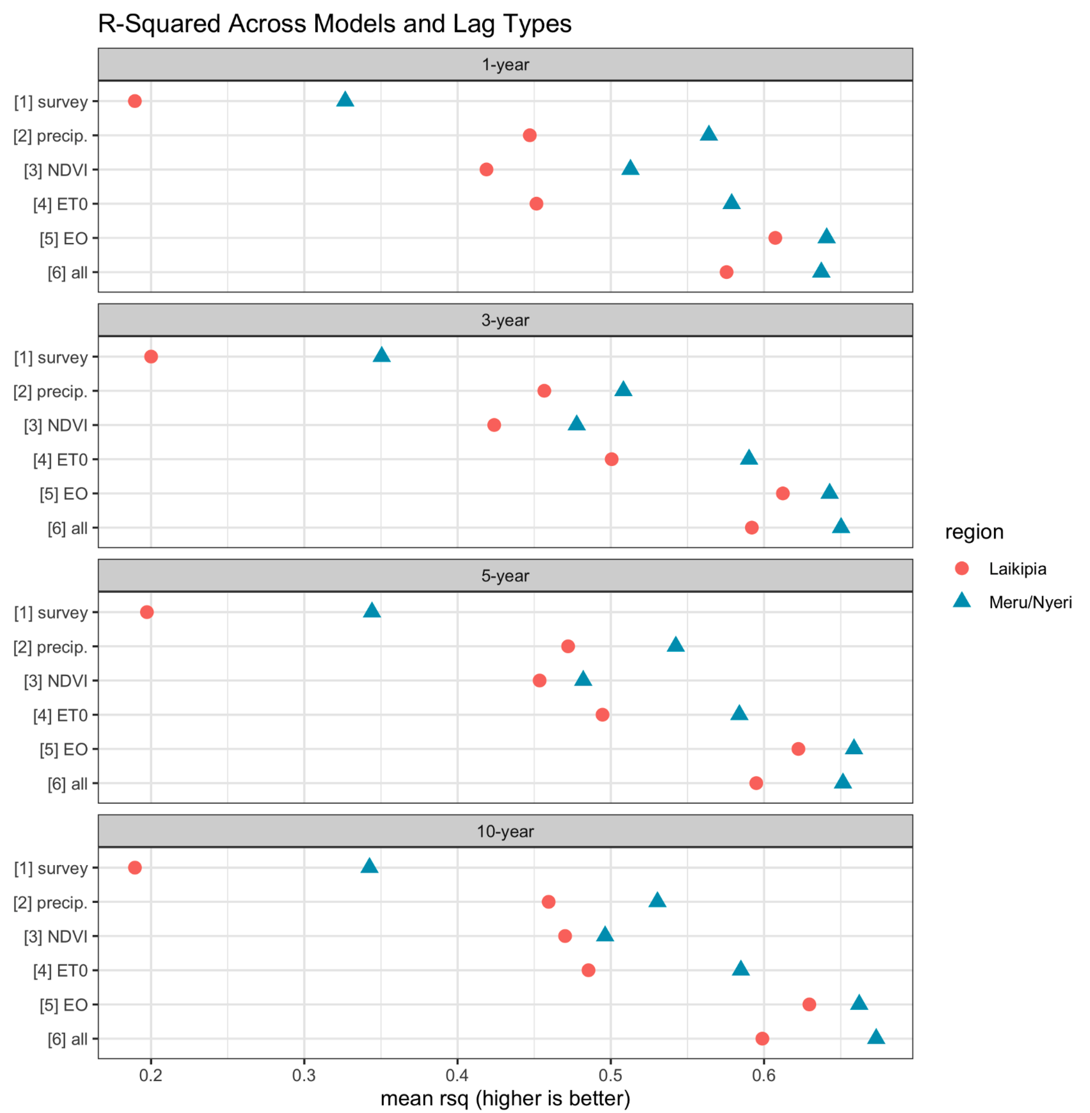

Fig. E.9. Random forest model results showing R-squared across models (rows within subplots, 1:6) and lag types (rows of subplots a:d). Results are subsetted by region where Laikipia county is demonstrated by the circle symbol; Meru and Nyeri counties are demonstrated by triangles.

Appendix F. Model results for individual years

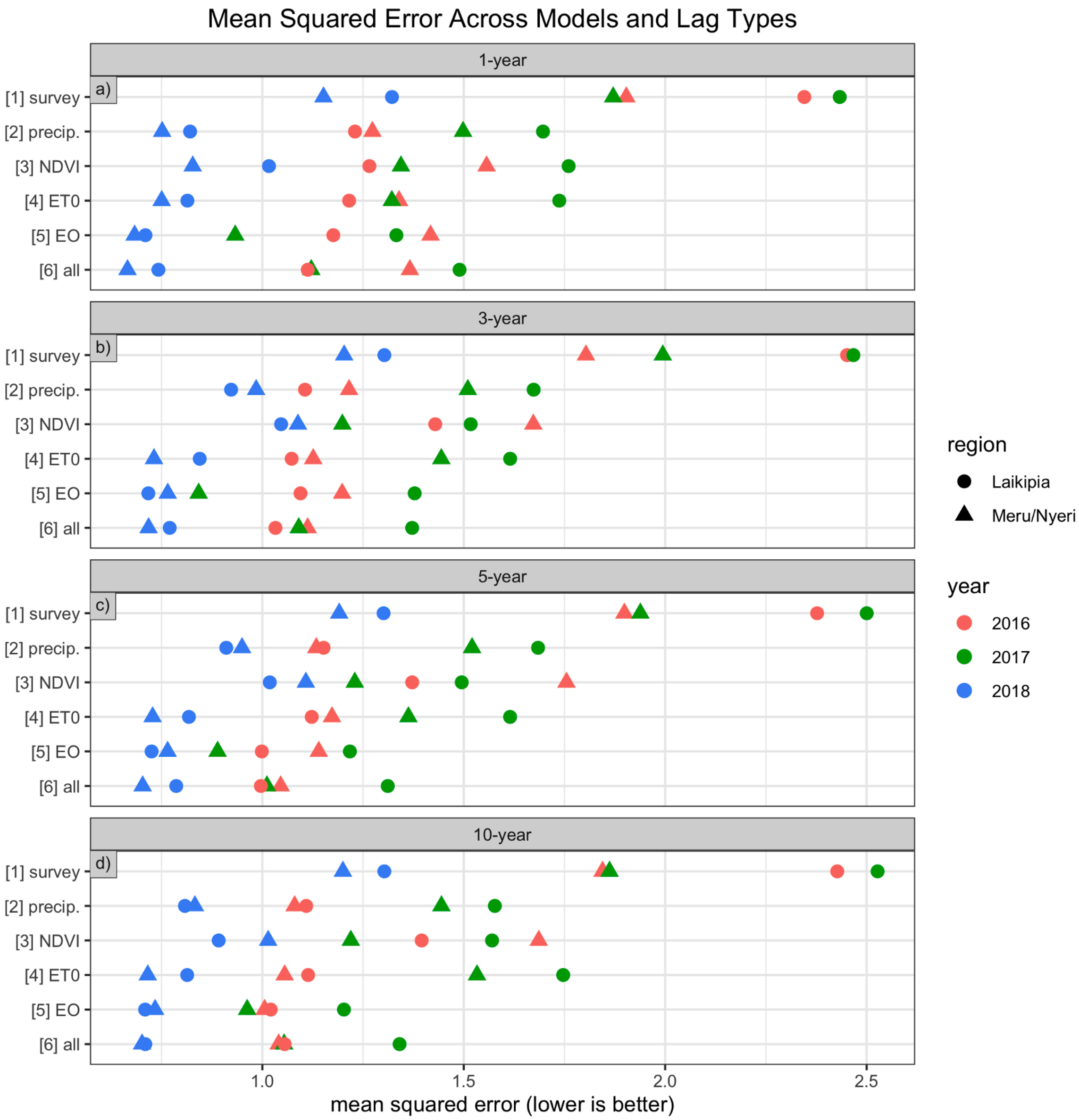

Fig. F.10. Random forest model results showing Mean Squared Error (MSE) across models (rows within subplots, 1:6) and lag types (rows of subplots a:d) by year. This graphic corresponds to Fig. 4 in the main text. Results are subsetted by region where Laikipia county is demonstrated by the circle symbol; Meru and Nyeri counties are demonstrat.ed by triangles.

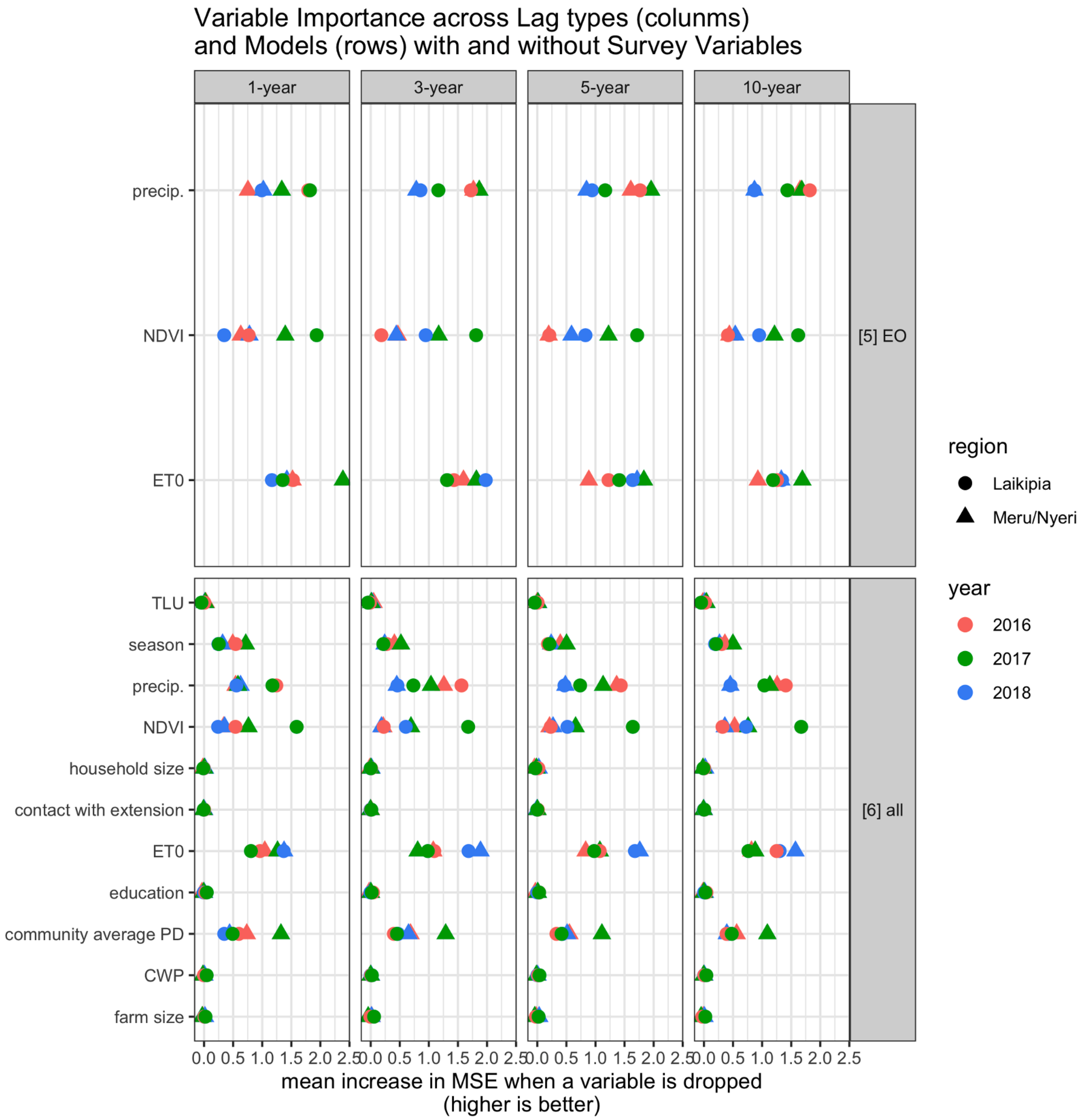

Fig. F.11. Random forest model results showing variable importance by year (corresponds to Fig. 5). Higher values in mean increase in MSE when a variable is dropped means higher importance of a given predictor variable. Lag types are presented as columns (a:h) and variables within models are presented as rows. Results are subsetted by region where Laikipia county is a circle symbol; Meru and Nyeri countie.s are triangles.

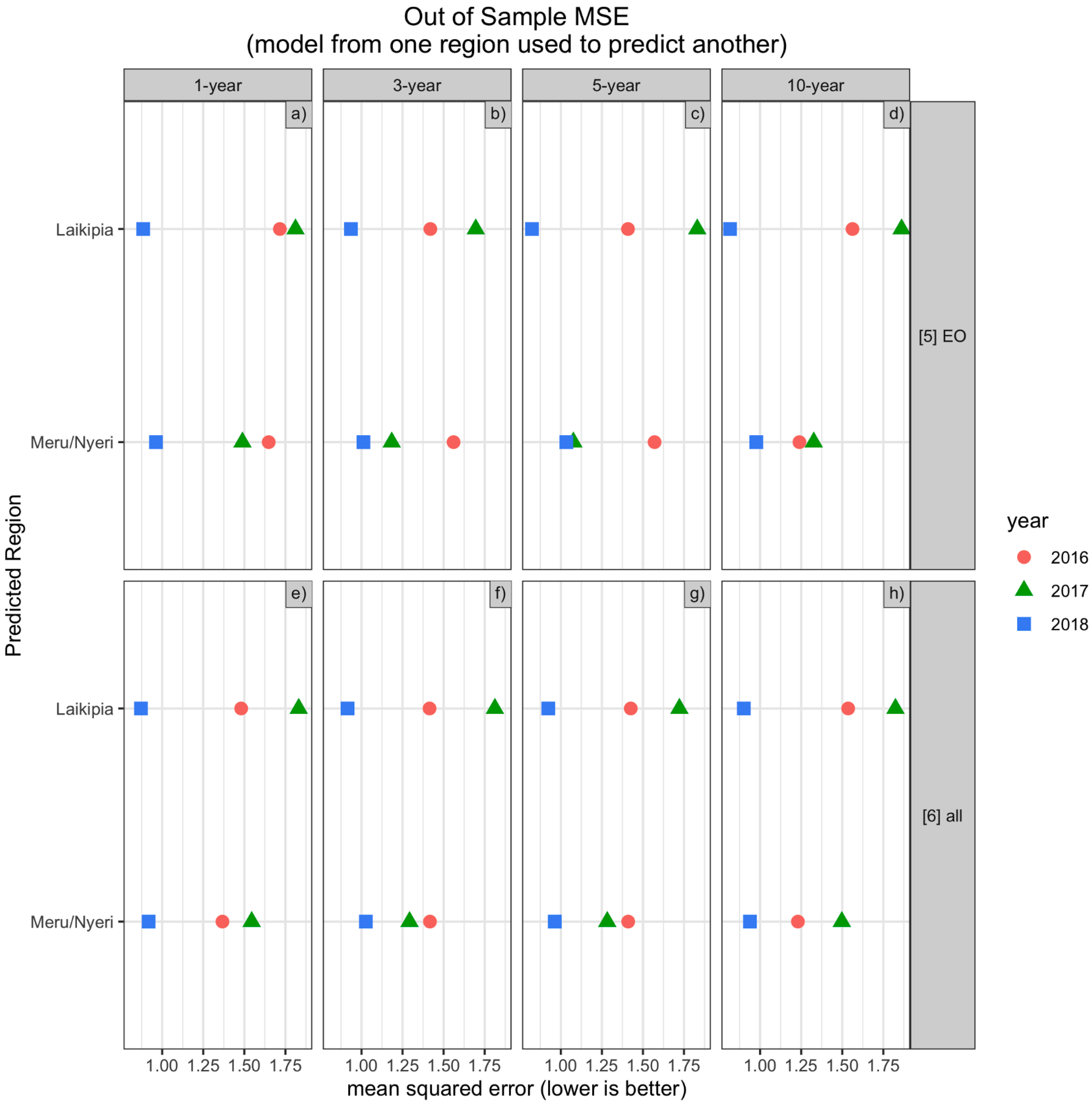

Fig. F.12. Out of Sample MSE results by year (corresponds to Fig. 6). Plot showing models 5-EO (top panel) and 6-all (bottom panel) Out of Sample MSE by region (rows) for the .SMS survey type.

References

Abrahão, G.M., Costa, M.H., 2018. Evolution of rain and photoperiod limitations on the soybean growing season in Brazil: The rise (and possible fall) of double-cropping systems. Agric. For. Meteorol. 256–257, 32–45.

AGRHYMET, Méthodologie de suivi des zones à risque, Technical Report, AGRHYMET FLASH, Bulletin de Suivi de la Campagne Agricole au Sahel, Niamey, Niger, 1996.

Belay, A., Recha, J.W., Woldeamanuel, T., Morton, J.F., 2017. Smallholder farmers' adaptation to climate change and determinants of their adaptation decisions in the central rift valley of ethiopia, Agriculture & Food. Security 6, 1–13.

Blandino, M., Scarpino, V., Giordano, D., Michael, S., Rudolf, K., Vanara, F., Reyneri Di Lagnasco, A., 2017. Others, Impact of sowing time, hybrid and environmental conditions on the contamination of maize by emerging mycotoxins and fungal metabolites. Italian J. Agronomy.

Bondeau, A., Smith, P.C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Müller, C., Reichstein, M., Smith, B., 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. Glob. Chang. Biol. 13, 679–706.

A. Borchers, E. Truex-Powell, S. Wallander, C. Nickerson, Multi-cropping practices: Recent trends in double-cropping, Technical Report, U.S. Department of Agriculture, 2014.

Bussmann, A., Elagib, N.A., Fayyad, M., Ribbe, L., 2016. Sowing date determinants for Sahelian rainfed agriculture in the context of agricultural policies and water management. Land use policy 52, 316–328.

D. Cammarano, B. Basso, L. Stefanova, P. Grace, Adapting wheat sowing dates to projected climate change in the Australian subtropics: Analysis of crop water use and yield, 2012.

Davenport, F.M., Shukla, S., Turner, W., Funk, C., Krell, N., Harrison, L., Husak, G., Lee, D., Peterson, S., 2021. Sending out an sos: using start of rainy season indicators for market price forecasting to support famine early warning. Environmental Research Letters 16, 084050.

J. Devèze, D. Halley des Fontaines, Le devenir des agricultures familiales des zones cotonnières africaines: une mutation à conduire avec tous les acteurs, A partir des cas du Bénin, du Burkina Faso, du Cameroun et du Mali. Agence Française de Développement, Paris, France (2005).

J.A. Dixon, D.P. Gibbon, A. Gulliver, Food and Agriculture Organization of the United Nations, Farming Systems and Poverty: Improving Farmers' Livelihoods in a Changing World, Food & Agriculture Org., 2001.

Dunning, C.M., Black, E.C.L., Allan, R.P., 2016. The onset and cessation of seasonal rainfall over Africa. J. Geophys. Res. 121, 11405–11424.

Eklundh, L., Jönsson, P., 2016. Timesat for processing time-series data from satellite sensors for land surface monitoring, in. In: Multitemporal Remote Sensing. Springer, pp. 177–194.

FEWS NET, Little rain forecast during the end of the Deyr/short rains season in the Horn of Africa, url:https://fews.net/east-africa/seasonal-monitor/december-15-2017, 2017. Accessed: 2021-2-17.

FEWS NET, 2019 Short Rains in east Africa among the wettest on historical record, url:https://fews.net/east-africa/special-report/january-29-2020, 2020. Accessed: 2020-1-29.

Fisher, M., Abate, T., Lunduka, R.W., Asnake, W., Alemayehu, Y., Madulu, R.B., 2015. Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa. Clim. Change 133, 283–299.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al., 2015. The climate hazards infrared precipitation with stations'a new environmental record for monitoring extremes. Scientific data 2, 1–21.

C. Funk, D. Pedreros, S. Nicholson, A. Hoell, D. Korecha, G. Galu, G. Artan, Z. Segele, A. Tadege, Z. Atheru, F. Teshome, K. Hailermariam, L. Harrison, C. Pomposi, Examining the potential contributions of extreme "western V" sea surface temperatures to the 2017 March–June east African drought, 2019.

C. Funk, F. Davenport, G. Eilerts, N. Nourey, G. Galu, Contrasting Kenyan resilience to drought: 2011 and 2017, 2018.

C. Funk, S. Shukla, W.M. Thiaw, J. Rowland, A. Hoell, A. McNally, G. Husak, N. Novella, M. Budde, C. Peters-Lidard, A. Adoum, G. Galu, D. Korecha, T. Magadzire, M. Rodriguez, M. Robjhon, E. Bekele, K. Arsenault, P. Peterson, L. Harrison, S. Fuhrman, F. Davenport, M. Landsfeld, D. Pedreros, J.P. Jacob, C. Reynolds, I. Becker-Reshef, J. Verdin, Recognizing the famine early warning systems network: Over 30 years of drought early warning science advances and partnerships promoting global food security, 2019.

Gichangi, E.M., Gatheru, M., Njiru, E.N., Mungube, E.O., Wambua, J.M., Wamuongo, J.W., 2015. Assessment of climate variability and change in semi-arid eastern Kenya. Clim. Change 130, 287–297.

Giroux, S.A., Kouper, I., Estes, L.D., Schumacher, J., Waldman, K., Greenshields, J.T., Dickinson, S.L., Caylor, K.K., Evans, T.P., 2019. A high-frequency mobile phone data collection approach for research in social-environmental systems: Applications in climate variability and food security in sub-Saharan Africa. Environ. Modelling Softw. 119, 57–69.

Giroux, S.A., McCord, P., Lopus, S., Gower, D., Dell'Angelo, J., Dickinson, S., Chen, X., Caylor, K.K., Evans, T.P., 2020. Environmental heterogeneity and commodity sharing in smallholder agroecosystems. PLoS One 15, e0228021.

Gower, D.B., Dell'Angelo, J., McCord, P.F., Caylor, K.K., Evans, T.P., 2016. Modeling ecohydrological dynamics of smallholder strategies for food production in dryland agricultural systems. Environ. Res. Lett. 11, 115005.

K. Guan, B. Sultan, M. Biasutti, C. Baron, D.B. Lobell, Assessing climate adaptation options and uncertainties for cereal systems in West Africa, 2017.

Guido, Z., Zimmer, A., Lopus, S., Hannah, C., Gower, D., Waldman, K., Krell, N., Sheffield, J., Caylor, K., Evans, T., 2020. Farmer forecasts: Impacts of seasonal rainfall expectations on agricultural decision-making in Sub-Saharan Africa. Climate Risk Management 30, 100247.

Guido, Z., Lopus, S., Waldman, K., Hannah, C., Zimmer, A., Krell, N., Knudson, C., Estes, L., Caylor, K., Evans, T., 2021. Perceived links between climate change and weather forecast accuracy: new barriers to tools for agricultural decision-making. Climatic Change 168, 1–20.

Hannah, C., Giroux, S., Krell, N., Lopus, S.E., McCann, L.E., Zimmer, A., Caylor, K.K., Evans, T.P., 2021. Has the vision of a gender quota rule been realized for community-based water management committees in Kenya? World Dev. 137, 105154.

Harris, D., Orr, A., 2014. Is rainfed agriculture really a pathway from poverty? Agricultural Systems 123, 84–96.

Harrison, L.S., 2018. In: Drivers of atmospheric evaporative demand during African droughts, 98th American Meteorological Society Annual Meeting.

Heumann, B.W., Seaquist, J., Eklundh, L., Jönsson, P., 2007. Avhrr derived phenological change in the sahel and soudan, africa, 1982–2005. Remote Sensing of Environment 108, 385–392.

Hobbins, M., Harrison, L., Blakeley, S., Dewes, C., Husak, G., Shukla, S., Jayanthi, H., McNally, A., Sarmiento, D., Verdin, J., 2018. Drought in Africa: Understanding and exploiting the demand perspective using a new evaporative demand reanalysis, 2018. American Geophysical Union, Fall Meeting.

Jahnke, H.E., 1982. Livestock production systems and livestock development in tropical Africa, volume 35. Kieler Wissenschaftsverlag Vauk.

Jain, M., Naeem, S., Orlove, B., Modi, V., DeFries, R.S., 2015. Understanding the causes and consequences of differential decision-making in adaptation research: Adapting to a delayed monsoon onset in Gujarat, India. Glob. Environ. Change 31, 98–109.

Jönsson, P., Eklundh, L., 2004. Timesat'a program for analyzing time-series of satellite sensor data. Computers Geosci. 30, 833–845.

Kilavi, M., MacLeod, D., Ambani, M., Robbins, J., Dankers, R., Graham, R., Titley, H., Salih, A.A.M., Todd, M.C., 2018. Extreme rainfall and flooding over Central Kenya including Nairobi city during the Long-Rains season 2018: Causes, predictability, and potential for early warning and actions. Atmosphere 9, 472.

Krell, N.T., Giroux, S.A., Guido, Z., Hannah, C., Lopus, S.E., Caylor, K.K., Evans, T.P., 2020. Smallholder farmers' use of mobile phone services in central Kenya. Climate Dev. 1–13.

Krell, N., Morgan, B., Gower, D., Caylor, K., 2021. Consequences of dryland maize planting decisions under increased seasonal rainfall variability. Water Resources Research e2020WR029362.

Laux, P., Kunstmann, H., Bárdossy, A., 2008. Predicting the regional onset of the rainy season in West Africa. Int. J. Climatol. 28, 329–342.

Laux, P., Jäckel, G., Tingem, R.M., Kunstmann, H., 2010. Impact of climate change on agricultural productivity under rainfed conditions in Cameroon—A method to improve attainable crop yields by planting date adaptations. Agric. For. Meteorol. 150, 1258–1271.

A. Liaw, M. Wiener, Classification and regression by randomforest, R News 2 (2002) 18–22. url:https://CRAN.R-project.org/doc/Rnews/.

B. Liebmann, I. Bladé, G.N. Kiladis, L.M.V. Carvalho, G.B. Senay, D. Allured, S. Leroux, C. Funk, Seasonality of African precipitation from 1996 to 2009, J. Clim. 25 (2012) 4304–4322.

Lopus, S., McCord, P., Gower, D., Evans, T., 2017. Drivers of farmer satisfaction with small-scale irrigation systems. Appl. Geogr. 89, 77–86.

MacLeod, D., Kilavi, M., Mwangi, E., Ambani, M., Robbins, J., Graham, R., Rowhani, P., Todd, M.C., 2020. Are Kenya Meteorological Department heavy rainfall advisories useful for forecast-based early action and early preparedness? Natural Hazards Earth System Sci. Discussions 1–27.

Marie, M., Yirga, F., Haile, M., Tquabo, F., 2020. Farmers' choices and factors affecting adoption of climate change adaptation strategies: evidence from northwestern ethiopia. Heliyon 6, e03867.

Marteau, R., Moron, V., Philippon, N., 2009. Spatial coherence of monsoon onset over Western and Central Sahel (1950–2000). J. Clim. 22, 1313–1324.

Marteau, R., Sultan, B., Moron, V., Alhassane, A., Baron, C., Traoré, S.B., 2011. The onset of the rainy season and farmers' sowing strategy for pearl millet cultivation in Southwest Niger. Agric. For. Meteorol. 151, 1356–1369.

McCord, P., Dell'Angelo, J., Gower, D., Caylor, K.K., Evans, T., 2017. Household-level heterogeneity of water resources within common-pool resource systems. Ecol. Soc. 22.

Mugalavai, E.M., Kipkorir, E.C., Raes, D., Rao, M.S., 2008. Analysis of rainfall onset, cessation and length of growing season for western Kenya. Agric. For. Meteorol. 148, 1123–1135.

M. Negash, Analysing the determinants of farmers' preference for adaptation strategies to climate change: evidence from north Shoa zone of Amhara region, Ethiopia, Ph.D. thesis, Doctoral dissertation, 2011.

Ngigi, S.N., Savenije, H.H.G., Gichuki, F.N., 2007. Land use changes and hydrological impacts related to up-scaling of rainwater harvesting and management in upper Ewaso Ng'iro river basin. Kenya.

Patt, A., Suarez, P., Gwata, C., 2005. Effects of seasonal climate forecasts and participatory workshops among subsistence farmers in Zimbabwe. Proc. National Acad. Sci. 102, 12623–12628.

Python Software Foundation, Python Language Reference, volume version 3.7, http://www.python.org, 2020.

Qualtrics, Qualtrics, 2019. url:https://www.qualtrics.com.
Rateb, A., Hermas, E., 2020. The 2018 long rainy season in Kenya: Hydrological changes and correlated land subsidence. Remote Sensing 12, 1390.
J. Rockstrom, N. Hatibu, T.Y. Owels, S.P. Wani, others, Managing water in rainfed agriculture, in: Water for Food, Water for Life: A Comprehensive Assessment of Water Management in Agriculture, International Water Management Institute, 2007, pp. 315–352.
Sacks, W.J., Deryng, D., Foley, J.A., Ramankutty, N., 2010. Crop planting dates: an analysis of global patterns. Glob. Ecol. Biogeogr. 19, 607–620.
Shongwe, P., Masuku, M.B., Manyatsi, A.M., 2014. Factors influencing the choice of climate change adaptation strategies by households: a case of mpolonjeni area development programme (adp) in swaziland. J. Agric. Stud. 2, 86–98.
Shukla, S., Husak, G., Turner, W., Davenport, F., Funk, C., Harrison, L., Krell, N., 2021. A slow rainy season onset is a reliable harbinger of drought in most food insecure regions in Sub-Saharan Africa. PLoS One 16, e0242883.
Srivastava, A.K., Mboh, C.M., Gaiser, T., Webber, H., Ewert, F., 2016. Effect of sowing date distributions on simulation of maize yields at regional scale–a case study in Central Ghana, West Africa. Agric. Syst. 147, 10–23.
Sultan, B., Baron, C., Dingkuhn, M., Sarr, B., Janicot, S., 2005. Agricultural impacts of large-scale variability of the West African monsoon. Agric. For. Meteorol. 128, 93–110.
Tang, J., Wang, J., Wang, E., Yu, Q., Yin, H., He, D., Pan, X., 2018. Identifying key meteorological factors to yield variation of potato and the optimal planting date in the agro-pastoral ecotone in North China. Agric. For. Meteorol. 256–257, 283–291.
T. Therneau, B. Atkinson, rpart: Recursive Partitioning and Regression Trees, 2019. url:https://CRAN.R-project.org/package=rpart, r package version 4.1-15.
P. Uhe, S. Philip, S. Kew, K. Shah, J. Kimutai, E. Mwangi, G.J. van Oldenborgh, R. Singh, J. Arrighi, E. Jjemba, H. Cullen, F. Otto, Attributing drivers of the 2016 Kenyan drought, 2018.
Van den Berg, J., 2013. Socio-economic factors affecting adoption of improved agricultural practices by small scale farmers in south africa. African J. Agric. Res. 8, 4490–4500.
Waha, K., Van Bussel, L.G.J., Müller, C., Bondeau, A., 2012. Climate-driven simulation of global crop sowing dates. Glob. Ecol. Biogeogr. 21, 247–259.
Wainwright, C.M., Finney, D.L., Kilavi, M., Black, E., Marsham, J.H., 2021. Extreme rainfall in east africa, october 2019–january 2020 and context under future climate change. Weather 76, 26–31.
K.B. Waldman, S.Z. Attari, D.B. Gower, S.A. Giroux, K.K. Caylor, T.P. Evans, The salience of climate change in farmer decision-making within smallholder semi-arid agroecosystems, 2019.
Waldman, K.B., Giroux, S., Blekking, J.P., Baylis, K., Evans, T.P., 2020. Smallholder food storage dynamics and resilience. Food Secur. 12, 7–20.
WMO, International Meteorological Vocabulary, Technical Report, World Meteorological Organization, Geneva, Switzerland, 1992.
M. Zhang, Monitoring agricultural behavior under climate change with cloud computing and satellite imagery, Ph.D. thesis, University of California, Berkeley, 2020.
Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using modis. Remote Sensing Environ. 84, 471–475.