Agricultural and Forest Meteorology 262 (2018) 147–156

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Comparing empirical and survey-based yield forecasts in a dryland agro-ecosystem

Yi Zhao[a,*], Noemi Vergopolan[b], Kathy Baylis[c], Jordan Blekking[d], Kelly Caylor[e,f], Tom Evans[g], Stacey Giroux[h], Justin Sheffield[i], Lyndon Estes[a]

[a] Graduate School of Geography, Clark University, Worcester, MA, 01610, USA
[b] Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, 08540, USA
[c] Department of Agricultural and Consumer Economics, University of Illinois-Urbana Champaign, Urbana, IL, 61801, USA
[d] Department of Geography, Indiana University, Bloomington, IN, 47405, USA
[e] Department of Geography, University of California Santa Barbara, Santa Barbara, CA, 93106, USA
[f] Bren School of Environmental Science and Management, University of California Santa Barbara, Santa Barbara, CA, 93106, USA
[g] School of Geography and Environment, University of Arizona, Tucson, AZ, 84721, USA
[h] Ostrom Workshop, Indiana University, Bloomington, IN, 47405, USA
[i] Department of Geography and Environment, University of Southampton, Southampton, SO17 1BJ, UK

ARTICLE INFO

Keywords:
Yield forecast
Empirical modeling

ABSTRACT

Accurate crop yield forecasts before harvest are crucial for providing early warning of agricultural losses, so that policy-makers can take steps to minimize hunger risk. Within-season surveys of farmers' end-of-season harvest expectations are one important method governments use to develop yield forecasts. Survey-based methods have two potential limitations whose effects are poorly understood. First, survey-based forecasts may be subject to errors and biases in the response data. For example, the weather variables that most impact yields may not be the same as those that farmers consider when shaping their yield expectations, thereby undermining forecast accuracy. Secondly, surveys are typically conducted late in the growing season, giving the government less advance notices of potential crop failures or low yields, and are costly to implement. Here we investigate these limitations within the context of Zambia's annual Crop Forecast Survey (CFS). Concerning the first limitation, we analyzed the differences between CFS-predicted yields and reported yields collected by Post Harvest Surveys, and found that excess rainfall during the planting stage was more important to the actual yield than to farmers' yield forecasts. For the second limitation, we evaluated whether a simple empirical yield forecast model could produce earlier and more accurate yield forecasts than the CFS. A random forest model using weather variables, soil texture, and soil pH as predictors were able to produce yield forecasts at the same or higher accuracy since the planting season.

1. Introduction

Agriculture is vital to Sub-Saharan African economies, averaging 17.2% of GDP in 2015 (Agriculture, 2018). Agriculture also provides the primary livelihood for a majority proportion of the region's populations (Burney et al., 2013; Rockström, 2000). Agriculture is expected to expand dramatically to meet the region's rapidly growing food demands (Searchinger et al., 2015), and small to medium farmers are expected to play a critical role in this expansion (Morris and Byerlee, 2009; Jayne et al., 2016).

This agricultural dependence makes the region's economies, as well as the food security of many households, highly vulnerable to climate risk. This risk is further exacerbated by two factors: 1) the vast majority of farms in Sub Saharan Africa (SSA) are rain-fed (Burney et al., 2013), with relatively limited prospects for significant irrigation expansion (You et al., 2011); 2) most crops are produced in savanna regions, which are characterized by strong rainfall seasonality that exhibits pronounced and increasing variability at both inter- and intra-annual scales (Falkenmark and Rockström, 2008; Verdin et al., 2005a; Gaughan and Waylen, 2012; Nicholson, 2000; Rowhani et al., 2011; Milgroom and Giller, 2013). Seasonality limits the number of crops that can be grown during the year, while variability can lead to crop failure during the rainy months when crops are grown. For example, within-season variability of precipitation intensity and frequency can have

* Corresponding author.
E-mail address: jeanzhaoyi@gmail.com (Y. Zhao).

https://doi.org/10.1016/j.agrformet.2018.06.024
Received 6 December 2017; Received in revised form 7 June 2018; Accepted 24 June 2018

large impacts on maize, rice and sorghum yields (Cabas et al., 2010; Guan et al., 2015; Rowhani et al., 2011).

In addition to rainfall variability, temperature extremes pose a substantial threat to food security. Previous work has shown that maize (Africa's most widely grown staple crop, accounting for 30% of planted area; Cairns et al., 2013) is globally sensitive to elevated temperatures (Lobell and Field, 2007), and African maize yields may be reduced by ≥1% for each degree day above 30 °C (Lobell et al., 2011).

Given SSA's strong dependence on agriculture, the region's susceptibility to climate variability and extremes, and the sensitivity of one of its main crops to these extremes, the ability to forecast the potential impacts of extreme weather on maize harvests is crucial to mitigating yield loss and the attendant risks to livelihoods. Two types of forecasts are important in this regard, the first being pre-season forecasts of expected growing season weather conditions, and the second being yield forecasts, which are made during the middle or second half of the growing season. Pre-season weather forecasts allow governments to inform farmers of potential extreme conditions, so that they can alter their management practices to mitigate losses (Ingram et al., 2002; Patt et al., 2005; Roncoli et al., 2009). Yield forecasts allow governments to anticipate changes in food supply before the end of the growing season, and thereby implement compensatory policies (e.g. increased trade or price stabilization mechanisms) that help to buffer production and price shocks to ensure food remains available and accessible (Dorosh et al., 2009; Mason and Myers, 2013; Minot, 2014). It follows that the accuracy of both weather and yield forecasts are essential for maintaining food security.

Two standard methods employed for yield forecasts are: (1) remote sensing-based yield prediction, (2) yield forecasts developed from representative surveys of farming households, in which farmers' report their expectations of the final harvest and their planted areas. Satellite-based methods are widely used throughout sub-Saharan Africa, and are typically based on drought and Vegetation Indices that are combined with models to forecast yields (Brown et al., 2007; Rojas et al., 2011; Verdin et al., 2005b). The challenge behind using such methods is that the accuracy of many of the necessary remote sensing inputs such as land cover maps can exacerbate uncertainties (Estes et al., 2018) inherent in forecasting.

Zambia is an example of a country that relies on the survey-based approach, one that also demonstrates how the results of this approach can have significant policy ramifications. During March of each year, Zambia's Central Statistical Office (a division of the Ministry of Agriculture and Livestock) conducts Crop Forecast Surveys (CFS). The CFS collects data on end of season harvests (measured in total kg or bags of crop) expected by representative farming households, as well as the total area that the farmer reported as being planted to the crop. These surveys are followed by year-end Post-Harvest Surveys (PHS) during which the actual harvest numbers are collected from the same households. The intention of the CFS and PHS is that the PHS is a revisit of the CFS households. However, enumerators do not always find the exact same households in 100% of cases and the exact proportion of the matches between CFS and PHS households is unknown. The CFS is used to construct national-level predictions of the season's total harvest, which is used by Zambia's Food Reserve Agency (FRA) to implement policies that aim to supply sufficient food and to maintain price stability. For example, during normal harvest years the FRA purchases maize from small-holders at above-market prices to encourage production. In contrast, during low harvest years, the FRA imports maize to sell to domestic millers at below-market rates (Mason and Myers, 2013).

Numerous potential sources of error in farmer-reported yield forecasts exist. Firstly, farmers may lack the skill or knowledge required to estimate yields before harvest, thereby introducing potential bias or error into the CFS. As one example, farmers may make their predictions by evaluating certain weather events that have less impact on final yield than other factors. Secondly, even if a farmer is skilled in predicting yield and bases this prediction on the right variables, events following the forecast date (e.g. pest, disease, or extreme weather events) may significantly impact yield.

Another potential shortcoming of a survey-based method such as the CFS is the relatively late timing of the resulting yield forecast, which reduces the lead time for policy-makers to respond. In addition, the survey-based approach is also expensive and time-consuming to implement.

Given its importance to the Zambian food security policy, both the potential errors and the late timing of the CFS could have substantial socio-economic consequences. As an example, an overestimate of total harvest from the CFS could lead to inadequate maize importation in a low-harvest year, leading to domestic food price spikes. Despite such potential consequences, the errors and biases in the CFS are not well-understood, nor are their potential causes. It is also unclear whether reliable yield forecasts can be obtained for less cost and earlier in the season. To our knowledge, no one has directly compared survey-based forecasts with other methods.

In this study, we set out to address these knowledge gaps. To examine the degree of error in the CFS and its potential causes, we first quantified the discrepancies between the estimated and actual yields from Zambia's crop forecast surveys (CFS) and post-harvest survey (PHS). We then used empirical models to identify the weather and soil variables that have the most significant influence in explaining variation in both the CFS and PHS yields, to evaluate whether farmers' predictions are informed by different variables than those most strongly correlated with final yield. We also examined whether post-survey weather impacts PHS yields, thereby contributing to CFS prediction error. Lastly, to investigate whether alternative methods could provide reliable forecasts at earlier points in the season, we evaluated whether an empirical agro-meteorological forecast model, trained with weather data from different points in the growing season, could predict end-of-season yields earlier and more effectively than CFS.

Our results provide valuable insights into yield forecasting methods that may be useful for helping to improve food security monitoring systems.

2. Methods

2.1. Study site

Zambia is a landlocked country in sub-Saharan Africa with a semi-arid subtropical climate (Fig. 1). Maize is the most widely planted and consumed staple crop in Zambia, and is primarily grown under rain-fed conditions (Mason and Myers, 2013) during a single growing season from November to May. Total rainfall ranges from approximately 700 mm in the South of the country to 1400 mm in the North (Fig. 1). The southwestern part of Zambia is especially dry.

In addition to these characteristics, Zambia is a useful case study because it is broadly representative of the larger region. From a socioeconomic perspective, Zambia's smallholder-driven agricultural sector is a primary source of livelihood for much of the population, as with many other SSA countries (Rockström, 2000). It is also situated within the belt of higher rainfall savannas that comprise much of Africa's landmass, one where smallholder farmers are expected to be a key part of agricultural development in the coming decades (Searchinger et al., 2015; Morris and Byerlee, 2009).

2.2. Data

2.2.1. Survey data

The Zambian government conducts a Crop Forecast Survey (CFS) in mid-March, two to three months after planting, and approximately two months before maize harvest. It then conducts a Post-Harvest Survey (PHS) in October or November following each growing season. We obtained CFS survey data for the time periods of 2001–2005, 2008–2014, and PHS survey data from 1991 to 2005, 2007, 2008, 2011