Agricultural and Forest Meteorology 359 (2024) 110213

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Contrasting performance of panel and time-series data models for subnational crop forecasting in Sub-Saharan Africa

Donghoon Lee [a,b,*], Frank Davenport [b], Shraddhanand Shukla [b], Greg Husak [b], Chris Funk [b], James Verdin [c]

[a] Department of Civil Engineering, University of Manitoba, Winnipeg, MB, Canada
[b] Climate Hazards Center, Department of Geography, University of California, Santa Barbara, CA, USA
[c] U.S. Agency for International Development, Washington, DC, United States

ARTICLE INFO

Keywords:
Earth observation
Panel data
Machine learning
Crop yield
Operational forecast
Sub-Saharan Africa

ABSTRACT

We comprehensively examine methodologies tailored for subnational crop yield and production forecasting by integrating Earth Observation (EO) datasets and advanced machine learning approaches. We scrutinized diverse input data types, cross-validation methods, and training durations, focusing on maize production and yield predictions in Burkina Faso and Somalia. Central to our analysis is the comparative assessment of using time-invariant features within a panel data (PD) model versus a time-series data (TD) model. The TD model performed well in predicting both production and yield, while the PD model offered comparable yield predictions. Time-invariant features such as livelihood zones, soil properties, and cropland extents enriched the spatial understanding of crop data, enhancing the R-squared by 0.09 (0.21) for production and 0.11 (0.03) for yield, with corresponding reductions in the Mean Absolute Percentage Error by 90 % (238 %) for production and 5 % (4 %) for yield in Burkina Faso (Somalia). While Burkina Faso's consistent crop data allowed for effective modeling with brief training, Somalia benefited from the adaptability of the PD model to crop statistics outliers, particularly with extended training in high-producing regions. The PD approach showed promise in addressing data gaps, although predicting crop productions for unobserved districts remained a challenge. Our findings highlight the harmonious integration of EO data and machine learning in the field of agricultural forecasting and emphasize the importance of region-specific methodologies, especially in the rapidly changing landscape of EO data convergence.

1. Introduction

Sub-Saharan Africa (SSA) faces escalating food insecurity challenges. In 2022 alone, 120 million individuals, representing 12 % of the population in SSA, faced acute food insecurity, experiencing significant malnutrition and struggling to meet minimum food consumption needs (FEWS NET, 2023; Mitra et al., 2022). Significant strides have been taken to bolster early warning systems for agricultural droughts and famine (Fritz et al., 2019; Funk et al., 2019). Recent advancements have spotlighted the role of Earth Observation (EO) technologies. With enhanced global availability and long-term records, such as global climate observations and simulations, EO has become a pivotal tool for forecasting agricultural trends and outcomes (Lee et al., 2022; McNally et al., 2019; Nakalembe et al., 2021; Nakalembe and Kerner, 2023; Shukla et al., 2021; Verdin et al., 2005). For instance, insights from long-term precipitation data allow for the prediction of both grain yield and market price (Anderson et al., 2024 in review; Davenport et al., 2021; Lala et al., 2021). Turner et al. (2022) use climate forecasts to derive the water requirement satisfaction index. Moreover, the immediate in-season prediction of crop production has grown indispensable for proactive famine monitoring, early warning measures, and strategic management in SSA (Nakalembe et al., 2021).

Crop yield prediction models typically fall into three categories: process-based crop modeling (Andreadis et al., 2017; Delincé, 2017), remote sensing based prediction (Lobell et al., 2015), or empirical or statistical modeling combining both methods (Davenport et al., 2019; Lee et al., 2022). Augmented by EO and Machine Learning (ML), empirical models, particularly statistical ones, have emerged as powerful tools for large-scale crop productivity forecasting (Schauberger et al., 2020; Van Klompenburg et al., 2020). Broadly, these

* Corresponding author at: E1-428 EITC, 15 Gillson Street, University of Manitoba, Winnipeg, MB, Canada.
E-mail address: Donghoon.Lee@umanitoba.ca (D. Lee).

https://doi.org/10.1016/j.agrformet.2024.110213
Received 3 November 2023; Received in revised form 4 August 2024; Accepted 3 September 2024
Available online 5 November 2024

statistical models include two approaches: time-series and panel data prediction, as traditionally categorized in the field of Economics and Social Sciences (Frees, 2004; Wooldridge, 2010). Specifically, the panel (or longitudinal) data tracks the same subjects over time, capturing multiple observations per subject. This approach accounts for heterogeneity between entities and over time, which cannot be detected with pure cross-sectional or time-series data (Baltagi, 2021; Wooldridge, 2010).

When examining regional agricultural statistics, panel data refers to an aggregated dataset comprising time-series of crop statistics across multiple administrative units. This type of data allows for a more nuanced analysis of agricultural trends and productivity by capturing dynamic changes and regional differences. For example, Hu and McAleer (2005) utilized a panel dataset of 30 Chinese provinces over a 7-year span (1991–1997) to estimate agricultural production efficiency. You et al. (2009) investigated the impact of climate on Chinese wheat yield growth using crop-specific panel data. Adah et al. (2017) analyzed cereal productivity in West Africa through panel data, demonstrating its effectiveness in capturing the complex dynamics of agricultural productivity across regions. Furthermore, Lu et al. (2020) found that non-irrigated crops in the United States are more sensitive to severe droughts compared to irrigated crops, using a panel data regression model.

For subnational agricultural prediction, a time-series data (hearafter TD) model is fitted to a single dataset within a given region, while a panel data (hereafter PD) model is fitted to data across multiple regions. Several studies have recently attempted to forecast subnational crop yields using the TD or PD models. For example, Lee et al. (2022) developed TD models to predict subnational maize yields for Burkina Faso, Malawi, Kenya and Somalia using four EO products and machine learning techniques. Laudien et al. (2022) use TD-based yield and harvested area forecast models to predict national level production of maize, sorghum, and millet in Burkina Faso. Laudien et al. (2020) also applied TD models to predict maize yields with climate drivers for Tanzania. Conradt et al. (2016) showed improved modeling of annual yield changes for winter wheat and silage maize in about 300 German counties by aggregating separately estimated time series models into cluster-based panel models. Davenport et al. (2019) developed panel models and evaluated the out-of-sample performance of end-of-season maize yield forecasts for Kenya and Somalia using five EO products.

Typically, TD models offer higher accuracy because they tailor predictions for a specific unit unless data is insufficient. On the other hand, PD models come with several advantages: they encompass more observations and variability and, hence, minimize the impacts of poor data quality, grant greater degrees of freedom, minimize multicollinearity, account for individual heterogeneity, and facilitate the easy identification and measurement of variable dynamics, as well as capturing dynamic shifts in cross-sectional units over time (Hsiao, 2007; Wooldridge, 2010) (Table 1). When it comes to modeling crop statistics, PD models further benefit from utilizing time-invariant (i.e., unchanging over time) data. This static data can shed light on the spatial heterogeneity of agricultural conditions, including aspects like land cover, cropping systems (e.g., rainfed vs. irrigation), soil properties, and more.

PD approaches can also be useful (or the only option) if the time series is short or uneven (an unbalanced panel) across administrative units, as is often the case in SSA. The PD approach, with its ability to fill data gaps, can estimate crop yields in regions or periods where data is scarce or absent.

Table 1
The comparative strengths and weaknesses of time-series and panel data models as delineated in the literature.

Type	Pros	Cons
Time-series data (TD) prediction	Higher accuracy for a specific unit	Vulnerability to data gaps and overfitting
Panel data (PD) prediction	Robust against data gaps and data quality issues; Incorporates diverse spatial predictors	Lower accuracy for a specific unit

Consequently, this study delves into the nuances of forecasting subnational crop productions and yields in SSA. Specifically, our objectives are to:

(1) Evaluate the efficacy of PD and TD models, and discern primary influential predictors,
(2) Determine whether incorporating time-invariant data improves the accuracies of PD and TD models for forecasting crop productions and yields compared to models that only include time-varying data,
(3) Assess the advantages of the PD model over the TD model and its potential to refine operational forecasts

To achieve this, we fit maize production and yield prediction models for Burkina Faso and Somalia, representing countries with abundant and sparse data, respectively. By employing diverse sampling methods, limiting the use of time-invariant data, and altering training periods, we aim to compare the predictive capacities of these modeling approaches and assess their performance across time and space.

2. Data

2.1. Crop data

In this study, we predict subnational maize production and yield, both critical for evaluating food insecurity in SSA. We sourced the subnational crop statistics from the FEWS NET Data Warehouse (FDW). This repository contains consolidated crop production, harvested area, and yield records from local governments, ensuring quality control. For Burkina Faso and Somalia, the original data sources are the Ministry of Agriculture of Burkina Faso and FAO's Food Security and Nutrition Analysis Unit, respectively. For Burkina Faso's main maize-growing season, the dataset comprises records from 45 administrative level-2 districts spanning 36 years, from 1985 to 2021. Similarly, for the Gu season of Somalia, the dataset encompasses 31 administrative level-2 districts with records spread over 26 years, from 1995 to 2021 (Table 2).

In both countries, the long-term trends are distinctly observable. While Burkina Faso's maize production, area, and yield display upward trajectories, Somalia's corresponding metrics reveal declining trends (Fig. 1). Fig. S1 depicts the temporal trends in maize production and yield across all districts, while Fig. S2 depicts the spatial distribution of the mean and coefficient of variation of maize production and yield.

While some studies employ filtering or detrending techniques to segregate these extended trends and assess interannual climate impacts, we incorporate these trends directly into our model, using a growing year variable to discern their contributions (Davenport et al., 2019). To ensure the subnational crop data aligns seamlessly with the latest administrative boundaries, we perform several calibration and data refinement steps, elaborated in Text S1. Including the crop statistics for Burkina Faso and Somalia used in this study, an open-access subnational crop statistics dataset for African countries has been developed for scientific analysis (Lee et al., 2024).

2.2. Input feature selection

To forecast crop productivity in SSA, a set of time-varying and time-invariant features were carefully selected from EO data and other relevant sources. The selection process was guided by the need to capture the essential climatic, soil, and agro-economic conditions that influence crop productivity. Below, we provide a detailed listing and justification for each category of features, supported by relevant literature. Also, specific details of the EO datasets described in Table 3.

Table 2
Subnational maize production and yield data used in this study.

Country	Growing season	Administrative level	Number of districts	Number of record years	Data source
Burkina Faso	Main season (June-August); harvest concludes in December	Level 2	45	37 years (1985–2021)	Ministry of Agriculture, Burkina Faso
Somalia	Gu season (March-May); harvest concludes in July	Level 2	31	26 years (1995–2020)	FAO's Food Security and Nutrition Analysis Unit

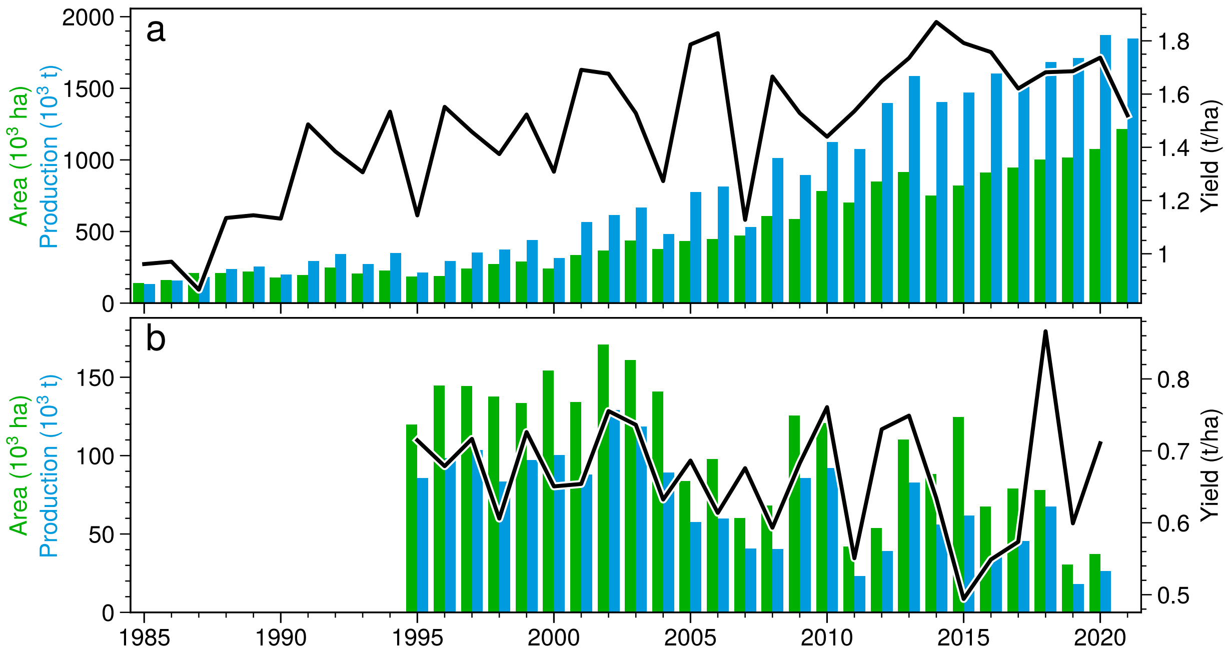

Fig. 1. National maize production, area, and yield for (a) Burkina Faso's main season and (b) Somalia's Gu season. We note that the combined national production and area figures might be marginally lower than the officially reported national values due to missing or excluded records at the administrative level-2 data.

Table 3
Overview of Earth Observation (EO) datasets used to predict maize production and yield. The "Features" column specifies the EO feature(s) derived from each dataset.

Category	Variable	Dataset	Spatial and temporal resolution, and latency	Features (references)
Time-varying features	Precipitation	CHIRPS (Climate Hazards Center InfraRed Precipitation with Station data) version 2 (Funk et al., 2015)	0.05° (~5 km), 1981–present, daily data, ~3 weeks latency; preliminary data is released 2 days after the end of each pentad (5 day period)	Precipitation, Dry days
	Temperature	NOAA CPC global daily maximum and minimum surface air temperature	0.5° (~50 km), 1979–present, daily data, ~1 day latency	Maximum temperature, Minimum temperature, Average temperature, Growing Degree Days (GDD), Killing Degree Days (KDD)
	Reference ET	NOAA Reference ET (ETo) Monitoring Dataset (uses MERRA2 atmospheric reanalysis) (Hobbins, 2016)	0.125° (~12 km), 1980–present, daily data, ~10 days latency	ETo for short reference crop (ETos), ETo for long reference crop (ETrs)
	Rootzone soil condition	Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) (McNally et al., 2017)	0.1° (~10 km), 1982-present, daily data, ~20 days latency	Rootzone soil moisture, Rootzone soil temperature
	NDVI	NOAA Climate Data Record Advanced Very High-Resolution Radiometer (AVHRR) version 5 NDVI	0.01° (~1 km), 1981–2019, daily data, currently non-operational	NDVI
		USGS eVMOD/eVIIRS NDVI	~375 m, 2002–present, dekadal, ~3 dekads latency (interim data with cloud masks are available)	
Time-invariant features	Cropland mask	IFPRI-IIASA cropland mask (Fritz et al., 2015)	0.0083°, static data	Cropland area (Cropland Area) and Proportion of cropland area to total area (Cropland Percent)
	Soil properties	Global Gridded Soil Information - SoilGrids (Hengl et al., 2014)	0.0083°, static data, four depth strata including D1 (0–5 cm), D2 (5–15 cm), D3 (15–30 cm), and D4 (30–60 cm).	Available soil water capacity (Soil WC), Soil organic carbon stock (Soil OCS), Soil pH index (Soil pH), Soil texture class (Soil Texture)
	Livelihood zone	FEWS NET Livelihood zone	Polygons of livelihood zones are rasterized and aggregated to administrative districts; 2014 (2015) Livelihood zone is used for Burkina Faso (Somalia)	Livelihood zone class

2.2.1. Time-varying features

Time-varying features capture the dynamic aspects of the environment that change over time and directly impact crop growth and yield (Table 3).

Precipitation Metrics: We use CHIRPS (Climate Hazards Center InfraRed Precipitation with Station data) precipitation data (Funk et al., 2015). Precipitation is essential for understanding water availability for crops, with studies demonstrating that precipitation patterns significantly influence crop productivity (Funk et al., 2008; Lobell et al., 2011). The number of dry days (days with less than 1 mm of precipitation) helps assess drought conditions and stress periods, which profoundly impact crop yields (Lee et al., 2022; Mishra and Singh, 2010). This metric is calculated based on CHIRPS daily precipitation.

Temperature Metrics: NOAA CPC global daily maximum and minimum surface air temperature data are used for temperature metrics, including maximum, minimum, and average temperatures, which are critical for crop development. High temperatures can lead to heat stress (Hatfield et al., 2011), while low temperatures can impact germination and growth (Porter and Gawith, 1999). Average temperature provides an overall indication of the conditions affecting growth cycles. Additionally, Growing Degree Days (GDD; sum of daily average temperatures over 10 °C) indicate crop growth potential (Mcmaster, 1997), while Killing Degree Days (KDD; sum of daily maximum temperatures over 30 °C) indicate potential damage to crops (Butler and Huybers, 2015).

Evapotranspiration Metrics: The NOAA Reference Evapotranspiration (ET) monitoring dataset, which includes solar radiation information (Hobbins, 2016), is used for ET metrics. ET for Short Reference Crop (ETos) serves as an indicator of water loss from a reference crop, critical for irrigation planning (Allen et al., 1998). ET for Long Reference Crop (ETrs) provides a comparative measure for a different crop reference (Jensen et al., 1990).

Rootzone Soil Metrics: Rootzone soil moisture and temperature are critical for understanding water availability and temperature effects in the root zone, which directly influence crop uptake and growth (Beauchamp and Lathwell, 1967). These metrics are sourced from the Famine Early Warning Systems Network Land Data Assimilation System (FLDAS) (McNally et al., 2017).

Vegetative Condition: We use the Normalized Difference Vegetation Index (NDVI), a key indicator of vegetation health and biomass (Anyamba and Tucker, 2012). We blended two NDVI datasets to cover the entire time period, including NOAA Climate Data Record Advanced Very High-Resolution Radiometer (AVHRR) NDVI version 5 for 1982–2002 and USGS EROS eVMOD/eVIIRS for 2002–2021. The eVMOD NDVI is a calibrated version of USGS EROS MODIS (eMODIS) NDVI, matched with eVIIRS NDVI using a geometric mean regression to consistently monitor drought-induced vegetative shifts (Skakun et al., 2018). The specific blending procedure is described in Text S2.

Using the International Institute for Applied Systems Analysis (IIASA) and International Food Policy Research Institute (IFPRI) cropland mask (Fritz et al., 2015), we spatially aggregated climate and vegetative conditions over cropland areas. We then consolidated the daily or dekadal data into monthly district-level values, using summation for precipitation, reference ET, and rootzone soil moisture features, GDD and KDD, averaging for other temperature features, and taking the maximum for NDVI. While alternatives to these datasets exist, our selection prioritizes their fine resolution for the SSA region and timely operational availability (Lee et al., 2022).

2.2.2. Time-invariant features

Time-invariant features represent the static characteristics of the environment that do not change significantly over time but are vital for understanding the baseline conditions of the agricultural landscape. Here, we incorporate three time-invariant datasets: cropland mask, soil attributes, and livelihood zones. The applicability of these spatial datasets can vary based on the specifics of each country. It's important to mention that we did not account for other critical aspects like fertilization, irrigation, and field management, primarily due to their inconsistent availability across both time and space.

Cropland Mask: Using the IIASA-IFPRI cropland mask, we quantified the actual cropland area and its relative proportion within the total land area of each district. The cropland area and the proportion of cropland area to total area provide insights into the agricultural intensity of the region. Larger cropland areas can indicate higher potential yields (Monfreda et al., 2008), while the proportion of cropland area to total area reflects the relative extent of cropland, indicating agricultural intensity (Fritz et al., 2015).

Soil Attributes: We obtained soil property data from the International Soil Reference and Information Centre (ISRIC) - World Soil Information's SoilGrids1km platform (Hengl et al., 2014). This includes metrics like available soil water capacity, soil organic carbon stock, soil pH measured in water solution, and soil texture classifications, which are critical for crop growth. The soil's ability to retain water (Rawls et al., 1982), soil fertility and health (Lal, 2004), nutrient availability, and microbial activity (Brady and Weil, 2008), and water retention and drainage properties (Hillel, 1982) all play essential roles in crop sustainability and productivity. The same crop mask is also applied to extract soil properties of croplands at the district level.

Livelihood Zones: We incorporated data from the FEWS NET's Livelihood Zone (LHZ). The LHZ class represents socio-economic classifications affecting agricultural practices and resilience strategies. The LHZ delineates regions with similar food and income sources, market access, local topography, environmental conditions, and farming practices, such as agropastoral or irrigated methods (FEWS NET, 2011). As LHZ typically spans multiple districts, we determined the predominant LHZ category on cropland in each district.

3. Methods

The TD model pertains to time-series data for a specific administrative unit, while the PD model captures multiple units concurrently, proving useful in large-scale analysis. One of the benefits of the PD model is that its parameters are tuned for multiple units, identifying significant factors across the entire region, such as rainfall patterns and soil quality, thereby aiding decision-making on food security assessments.

As panel data is utilized, model performance also depends on how the data is trained and tested, typically through cross-validation over time and space. This is crucial for estimating crop statistics that are unavailable for certain regions or years, enabling effective data gap filling. This study investigates four cross-validation techniques: Random K-fold (RK), Leave-District-Out (LDO), Leave-Year-Out (LYO), and Time-series (TS) cross-validations, as illustrated in Fig. 2.

- Random K-fold (RK) cross-validation: This method utilizes a randomized selection of crop data, integrating both spatial and temporal dimensions. It's commonly employed in spatio-temporal machine learning models (Meyer et al., 2018; Wang et al., 2023). In our study, the number of years determines the K value, which ensures the similar ratio between training and test samples as other cross-validation techniques. The RK method is ideal for estimating missing crop data because it employs data from both the target district and year.
- Leave-District-Out (LDO) cross-validation: Using a leave-one-out cross-validation (LOOCV) in the spatial dimension, LDO retains all spatial and temporal data while focusing on a specific district. It's designed to predict data for districts with no existing records.
- Leave-Year-Out (LYO) cross-validation: Like LDO, this employs a LOOCV approach but in the temporal dimension, centering on a particular year. This method is intended to simulate either a retrospective analysis of past years or a forecast of future years.
- Time-series (TS) cross-validation: A specialized method, TS employs LOOCV on the temporal dimension while focusing on a designated

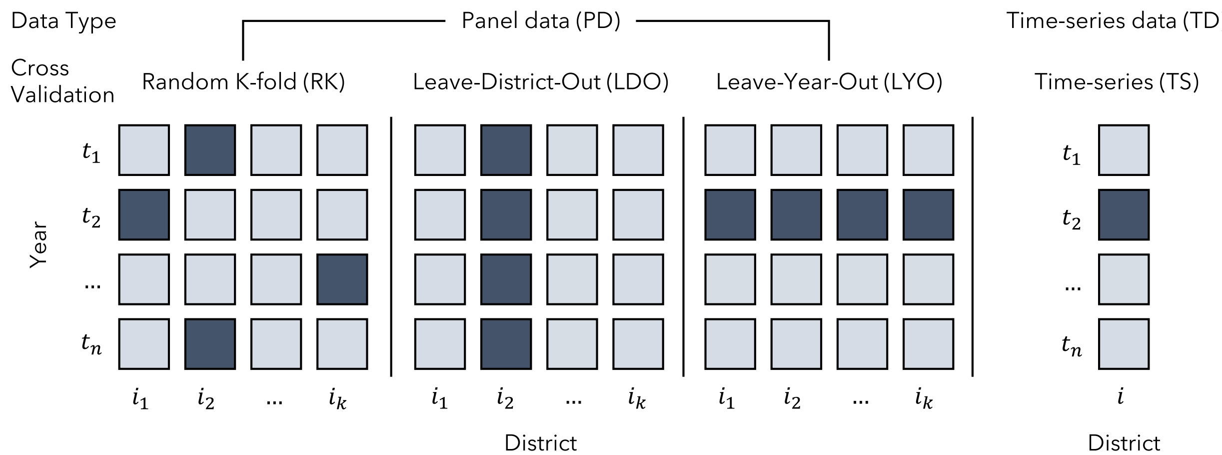

Fig. 2. An overview of the four cross-validation methods employed for crop production and yield prediction: Random K-fold (RK), Leave-District-Out (LDO), Leave-Year-Out (LYO), and Time-series (TS) cross-validations. Light-colored cells indicate training samples, while dark-colored cells indicate testing samples.

district. It is unique in that it only considers temporal variability of the district of interest.

Through the exploration and implementation of these cross-validation methods, we aim to conduct a thorough assessment and comparison of their performance in crop yield and production estimation, specifically focusing on prediction and the data gap filling.

Next, given the sporadic nature of crop statistics in SSA, we assess how data availability influences operational forecast precision. We developed LYO and TS models using datasets covering increments of 5, 10, 15, or 20 years, culminating in the most recent 5 years. In this framework, each year ($year_i$) of the last five years was predicted using data from all preceding years (from $year_0$ to $year_{i-1}$), reflecting a standard operational approach. We term this method 'operational prediction', whereas the approach that incorporates future years into the training set (as shown in Fig. 2) is termed 'cross-validated prediction'. Therefore, while cross-validated predictions utilize both past and prospective data, operational predictions solely depend on historical data for training.

Prediction accuracy is primarily assessed using the out-of-sample R-squared metric, complemented by Mean Absolute Percentage Error (MAPE). To evaluate the influence of time-invariant features, we made predictions both including and excluding them, contrasting their outcomes with predictions based exclusively on time-varying features. For clarity, when both time-varying and time-invariant features are incorporated together, we refer to this as combined features. In both cases, we integrated the variables "year" and "district" into the PD model. While "year" was treated as a continuous variable to identify long-term trends, "district" was categorized to highlight regional disparities, such as districts with notably high or low production (Davenport et al., 2019). Each district was denoted by a unique integer. Given that time-invariant features lack temporal variability and thus cannot be utilized in the TD model, predictions using either time-varying or combined features will yield equivalent results. Overall, we employed 14 time-varying features and 26 combined features (see Table 3). Time-varying features showcase unique temporal patterns throughout a crop calendar, with pronounced signals detected in early precipitation and late-stage NDVI (Lee et al., 2022). Drawing insights from the crop calendar, monthly fluctuations, and correlations with end-season maize production (as depicted in Figs. S3 and S4), we utilized seasonal time-varying features from main growing periods – July to September for Burkina Faso and March to May for Somalia – as predictors. For NDVI, we applied a one-month lag relative to the main growing season, accounting for its peak correlation with crop data occurring one month later than the rainy season (see Figs. S3 and S4). For consistency, we used the same seasonal predictors across all districts to streamline our experimental approach.

For predictive modeling, we employed the XGBoost (eXtreme Gradient Boosting) regression model (Chen and Guestrin, 2016). Hyperparameter tuning was conducted using Bayesian optimization search cross-validation from the `scikit-optimize` Python package. We assessed feature importance (FI), which indicate the contribution of each feature to the model's predictions, by applying the permutation feature importance technique to predictions made using all data points. This technique randomizes each feature in the test set and assesses the subsequent impact on model performance (Breiman, 2001).

4. Results

4.1. Significance of predictors via feature importance

Feature importance was initially evaluated based on predictions made using combined features (Fig. 3). Consistently across different evaluations, the "year" variable ranked highly, highlighting the model's emphasis on temporal trends. This is particularly pronounced in Burkina Faso's data, where a strong upward trend outweighs other spatial and temporal fluctuations, as visualized in Fig. 1a.

Although the combined features primarily augment the time-varying features with time-invariant data, they frequently offer superior performance. Notable time-invariant features that stand out include livelihood zone, soil pH, soil texture, soil organic carbon stock, and cropland percentage. These results suggest that the PD model leverages spatial attributes to categorize or determine crop productivity levels, such as differentiating between high and low production districts. For instance, the livelihood zone in Burkina Faso, which encapsulates regional agricultural activities like cereals, livestock, and pastoralism, is a top indicator for maize production. This is due to its ability to provide spatial insights on production levels.

In Somalia, features like cropland percentage and soil attributes, including texture and organic carbon stock, rank prominently (Fig. 3cd). The importance of these features can be attributed to the unique characteristics of the region, such as sparse cropland in lower production districts and the distinct distribution of soil properties (seen in S2). The model often flags soil texture and root zone soil attributes as pivotal, primarily because they critically influence crop yields and nutrient responses (Maman et al., 2018). Following these top-tier time-invariant features, time-varying features like NDVI, precipitation, ETo, KDD, soil temperature, and maximum temperature become relevant, as illustrated in Fig. 3.

The "year" and "district" variables emerge as the most important features when the model is confined to time-varying features (Fig. S5). They are closely followed by features such as NDVI, precipitation, soil moisture, and a variety of temperature metrics. Remarkably, while the

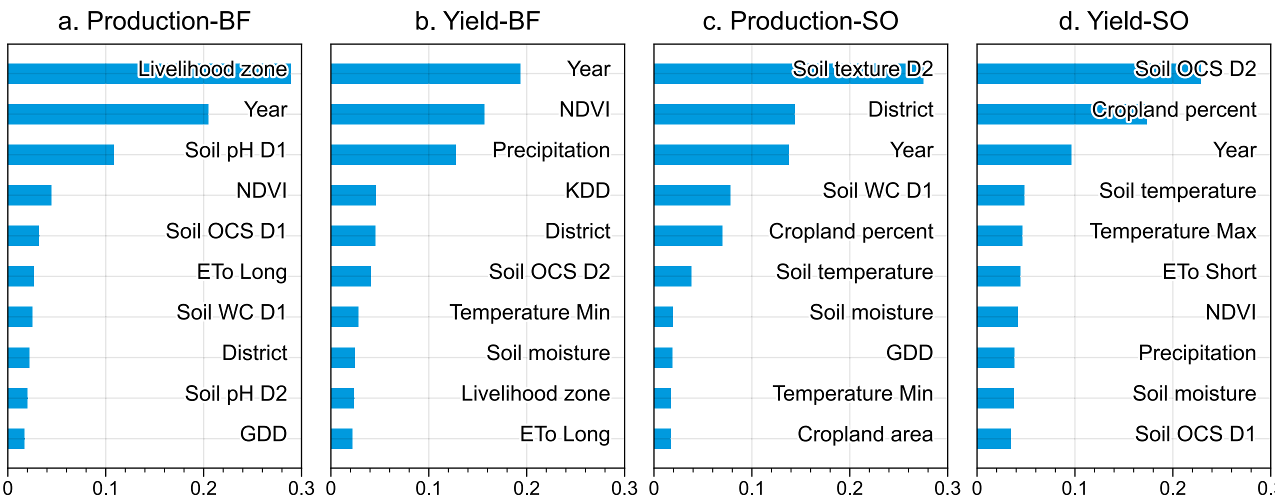

Fig. 3. Top 10 feature importance for (a) maize production and (b) maize yield of Burkina Faso, and (c) maize production and (d) maize yield of Somalia. Predictions are based on combined features and all data points.

Table 4
Accuracy metrics for PD model predictions of maize production and yields in Burkina Faso (BF) and Somalia (SO) with time-varying features only and combined features. The PD model is trained and tested on entire panel data, with a random split of 80 % for training and 20 % for testing.

Metric	Production-BF		Yield-BF		Production-SO		Yield-SO	
	Time-varying	Combined	Time-varying	Combined	Time-varying	Combined	Time-varying	Combined
R-squared	0.81	0.90	0.53	0.64	0.52	0.73	0.60	0.63
MAPE	190 %	100 %	33 %	28 %	609 %	371 %	53 %	49 %

"district" frequently stands out as the primary feature, it is superseded by "NDVI" for maize yield prediction in Burkina Faso.

Overall, incorporating time-invariant features improves the accuracy of predictions, with an increase in R-squared values ranging from 0.03 to 0.21 and MAPE values ranging from 5 % to 238 % when compared to using only time-varying features (Table 4). This emphasizes the crucial importance of spatial characteristics in the panel model, especially for a comprehensive grasp of the spatial distribution in crop data.

4.2. *Cross-validated accuracy with different cross-validation methods and input data*

The prediction accuracy for various cross-validation methods and input data types are presented in Fig. 4. In general, our predictive framework exhibits high accuracy for maize production with R-squared values between 0.7 and 0.85. In contrast, maize yield predictions offer moderate accuracy, registering R-squared values from 0.45 to 0.5 in both countries. This suggests that while crop production's temporal trends and spatial variations can be more readily associated with the predictors, the same is more challenging for crop yield. For instance, the model more readily distinguishes between regions of high and low maize production than it does for maize yield. Burkina Faso exhibits a more pronounced accuracy disparity between production and yield compared to Somalia. Comparing the average accuracy of all cross-validation methods across countries, maize production predictions are marginally more accurate in Burkina Faso, whereas maize yield predictions are slightly more precise in Somalia.

As anticipated, of the four cross-validation methods, RK—enriched with both spatial and temporal information via the combined features—yields the highest accuracy. TS also exhibits commendable accuracy, signifying the robustness of a model tailored for specific districts. Indeed, when predictions are based exclusively on time-varying features, the TS method frequently outperforms RK (Fig. 4). Pinpointing the primary influential features from district-specific models is challenging; however, the year, precipitation, NDVI, soil moisture, and ETo short grass consistently emerge with high feature importance rankings (Fig. 3). Surprisingly, the LYO method yields accuracy levels that are not only comparable but occasionally even surpass those of the TS method, particularly when time-invariant features are incorporated.

In predictions utilizing RK and LYO, incorporating time-invariant features bolsters prediction accuracy by R-squared values ranging from 0.05 to 0.15. This underscores the importance of comprehensively understanding a target district's spatial attributes—such as the livelihood zone or the proportion of cropland—as shown in Fig. 3, to more accurately attribute temporal variability to predictors. However, the accuracy of LDO predictions is diminished when time-invariant features are employed. Since LDO draws on spatial information from other districts for training, this can skew predictions if the target district vastly differs in certain attributes, such as production levels. These spatial mismatches in LDO predictions are more detrimental to production predictions than yield predictions due to the greater inherent spatial variability of production data. This is particularly true for Somalia's production predictions, which grapple with pronounced production disparities across districts, even culminating in entirely errant predictions as evidenced by negative accuracy values. Lastly, the improvements of time-invariant features in LYO prediction of maize yield in Somalia appear to be minimal (Fig. 4).

4.3. *Operational accuracy of PD and TD models predictions with different training periods*

For our operational evaluations, we narrowed our focus to the LYO-based PD model and TD (TS) model. To assess how prediction performances vary with the length of training data, we developed models using the combined features over varying training durations: 5, 10, 15, 20, and 25 years, up until the last 5 years of data for each country. Then, we measured the out-of-sample accuracy metrics for the predictions across all districts over the last 5-year period, where predictions were made using all preceding years as is standard in operational settings.

In Burkina Faso, the accuracy remains relatively consistent across

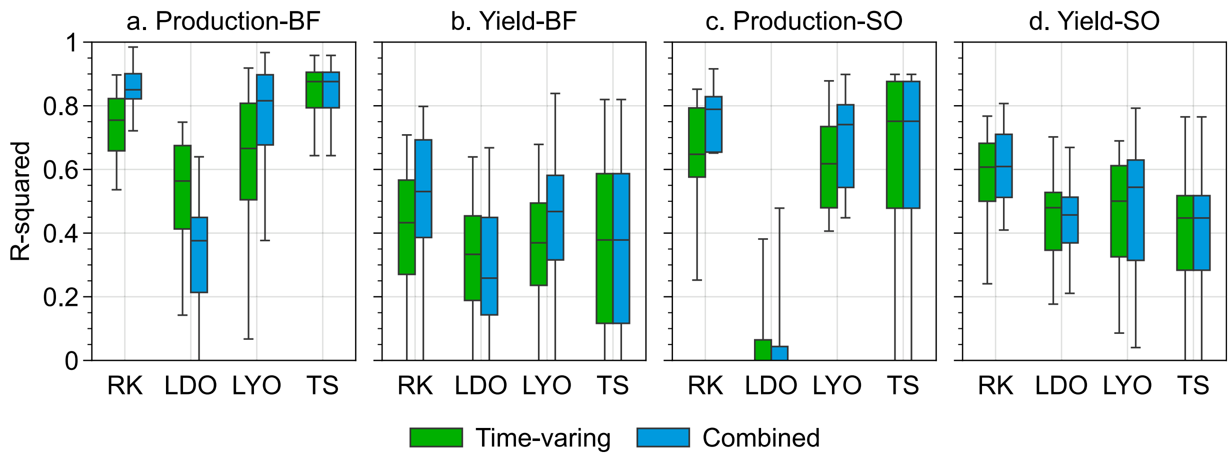

Fig. 4. Comparative accuracy in predictions based on various cross-validation methods and input data types for (a) maize production and (b) maize yield of Burkina Faso, and (c) maize production and (d) maize yield of Somalia. The boxplots depict the annual distribution of R-squared values.

various training periods and between PD and TD model predictions and for production as well as yield forecasting. Notably, predictions derived from just 5 years of data are almost as accurate as those based on longer training datasets. This suggests that the 5-year data sufficiently captures most influential predictors, such as temporal trends and livelihood zones (as highlighted in Fig. 3′s feature importance), allowing for precision even without longer historical data. Moreover, the consistency in trends and spatial variability over the 25-year span (referenced in Figs. 1 and S1) likely contributes to this stability in predictive accuracy over all training periods.

Furthermore, in Somalia, PD model outperforms TD model in predicting both maize production and yield. TD model predictions show consistent performance with the initial 5-year training data but presents a decline in accuracy as the training period extends. This drop in accuracy becomes particularly pronounced when transitioning from 20 to 25 years of training data. In contrast, PD model's accuracy shows an upward trajectory with longer training spans. Such variation can be attributed to prominent fluctuations during 1995–2005, which were characterized by elevated maize production in a few districts and slightly positive trends that diverged from subsequent years (as depicted in Figs. 1 and S1). As TD models don't inherently discriminate between high or low production areas and lean heavily on temporal variability, abnormal training data can skew predictions. However, PD's steadily increasing accuracy highlights its prowess in balancing both temporal and spatial variances, making it more adept with extensive data. Despite the different testing periods, the TD model prediction of Somalia's production has a lower operational accuracy than cross-validated accuracy, whereas PD model predictions have comparable operational and cross-validated accuracy (Figs. 4 and 5). This also implies that the PD model forecast is less susceptible to error from an operational standpoint.

Similar trends in operational accuracy is also observed in the MAPE metric (Fig. S6). Due to its exclusive emphasis on percentage differences, MAPE generates a comparatively higher level of error in comparison to R-squared metric, particularly with regard to production predictions. This is especially substantial in Somalia, where 70 % of total production is attributed to only seven primary districts. Overall, similar to the R-squared metric, there are marginal differences in the MAPE scores between PD and TD model predictions across various training periods. Similarly, for maize production predictions in Somalia, the PD model shows 's MAPE shows marked improvements, whereas the TD model's MAPE presents a decline in accuracy (Fig. S6).

The spatial variance in operational accuracy between PD and TD models is visually illustrated in Fig. 6. Overall, TD model is better for predicting maize production and PD model is better for predicting maize yields. This result aligns with the spatial distribution observed in cross-validated accuracy assessments (as in Fig. S7), though the differences in accuracy are not substantial. These patterns also resonate in regions with high maize production – areas crucial for gauging food security. Nonetheless, for maize production predictions in Somalia, PD model emerges as the more proficient method for major producing regions, especially given its adeptness in considering the irregular production statistics from 1995 to 2005 using spatial context.

5. Discussion

Our research conducted an in-depth exploration of methodologies tailored for predicting subnational crop productions and yields in Burkina Faso and Somalia. We primarily assessed the efficacy of merging time-varying Earth Observation (EO) features with time-invariant features in a panel data (PD) model, contrasting it with a time-series data (TD) model. Initially, we analyzed influential predictors using feature importance, both including and excluding time-invariant features, to gauge their significance. We then constructed maize production and yield prediction models using three PD-based and one TD cross-validation techniques: Random K-fold (RK), Leave-District-Out (LDO), Leave-Year-Out (LYO), and Time-series (TS) cross-validations. Additionally, we established models over different training durations to understand the impact of data availability on forecast accuracy.

Of the four cross-validation methods evaluated, RK emerged as the most accurate and reliable across all scenarios. This highlights its potential to fill data gaps in crop statistics by combining spatial and temporal variability. While TS trailed closely behind RK in terms of accuracy, it did not surpass it. However, there were instances, such as in maize yield predictions, where LYO sometimes outperformed TS. Generally, when accuracy is the most important factor, TS is the preferred method for predicting production, whereas LYO excels at predicting yield. With our model settings, LDO predictions exhibited weak performance for production forecasting, failing to predict production for Somalia, while showing only moderate success in yield predictions.

We observed that certain time-invariant features, such as livelihood zone, soil properties, and proportion of cropland area, effectively account for crop data characteristics. Indeed, the most dominant features were identified as non-EO products, including either a time-invariant feature or time itself (Fig. 2). This highlights the considerable impacts of arable land quality and agroecological information on crop production modeling, overshadowing some of the agroclimatic EO predictors. Among the EO predictors, NDVI and temperature-based variables are prominent, with precipitation behind those variables. This insight is valuable when applying our methods to regions with similar agroclimatic environments. For such regions, PD model emerges as a

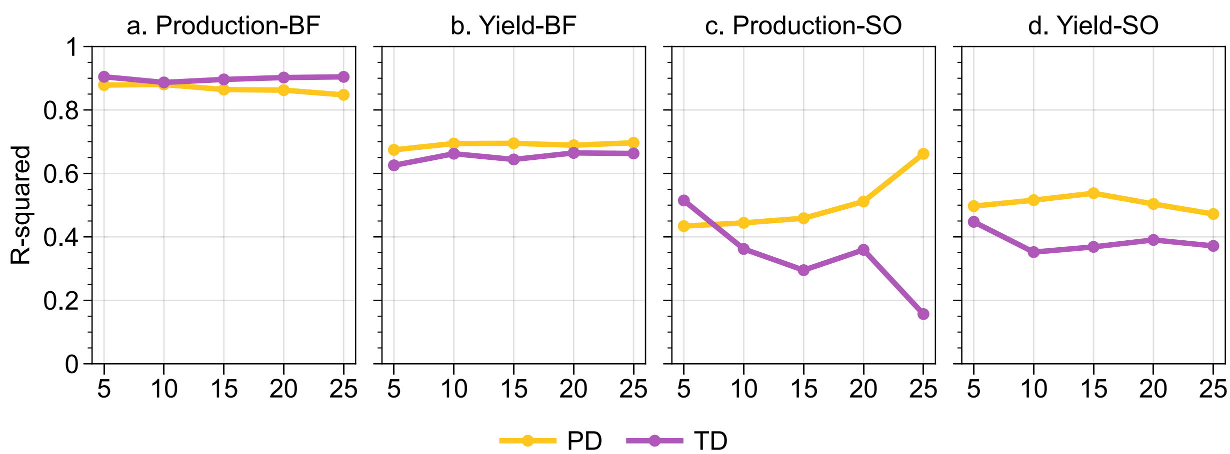

Fig. 5. Comparative operational accuracy (R-squared) between PD (LYO-based) and TD (TS-based) model predictions with varying training periods for (a) maize production and (b) maize yield of Burkina Faso, and (c) maize production and (d) maize yield of Somalia.

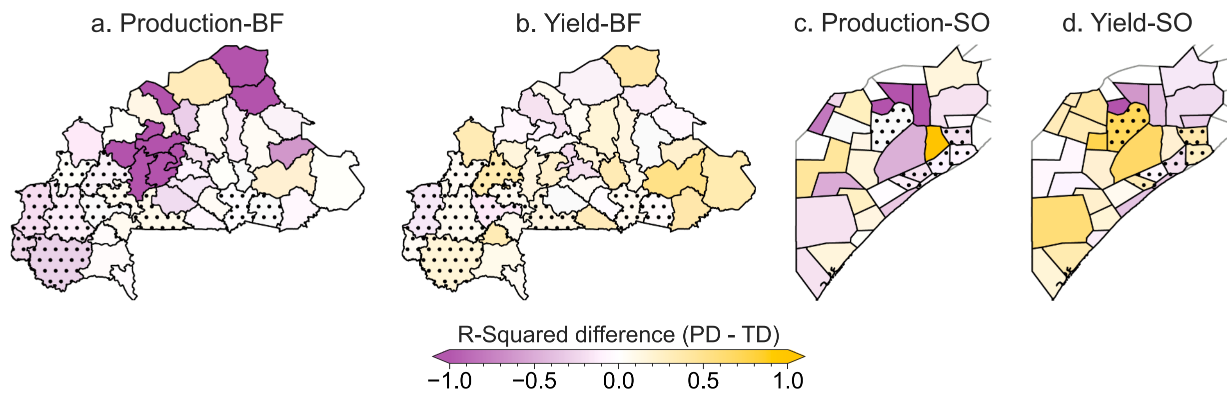

Fig. 6. Comparative operational accuracy between PD (LYO-based) and TD (TS-based) model predictions with a 20-year training duration for (a) maize production and (b) maize yield of Burkina Faso, and (c) maize production and (d) maize yield of Somalia. Positive values indicate superior performance by PD model relative to TD model. The dotted regions are the primary maize-producing regions, accounting for 70 % of each country's total production.

potentially superior choice. Also, our results indicate that time-invariant data can offer more than just capturing spatial variability. It can also enhance the temporal adaptability of the model to the target district, such as high producing regions in Somalia. When using all data points, including time-invariant features improves prediction accuracy by 0.05 to 0.41 in R-squared and 29 % to 2 % in Mean Absolute Percentage Error values compared to using only time-varying EO features.

In Burkina Faso, both production and yield predictions displayed stable operational accuracy across different training periods. This suggests that even a concise training duration can adequately capture spatial and temporal variability, given the consistency of crop statistics over time in Burkina Faso. In contrast, for Somalia, the PD model exhibited a stronger ability to handle crop production outliers compared to the TD model, with accuracy generally increasing with longer training periods. In general, we did not identify a consistent trend linking accuracy to the length of data. This highlights the prevalent impact of inherent crop characteristics and underscores the necessity for conducting experiments that are tailored to individual countries.

Depending on the type of missing data—whether it be an entire district or specific years—tailored strategies and methods are required for filling in the gaps. Our findings indicate that filling in sporadic gaps in crop statistics is feasible, allowing the development of a PD model that incorporates both spatial and temporal information. While our method struggles to predict complete records for new districts, performance may be improved by clustering districts based on their agro-ecological characteristics before developing a PD model.

Although our findings demonstrated that PD approach has potential to improve operational crop productivity forecasting in terms of prediction accuracy at specific condition, we also demonstrated that those improvements can vary with different experimental setting (e.g., cross validation and input features) and characteristics of crop statistics (e.g., outliers). Furthermore, operational applicability in SSA might need throughout examination on how these improvement affects to accuracy of lower production regions that food insecure are prominent or major producing regions that can affect provincial crop price shock and food supply shortage. Since the performance of TD model is also comparable or superior to PD model, additional examination is required to decide how to consolidate those approaches. Blending methods that combine the advantages of both PD and TD, such as spatially integrated predictions, have the potential to improve the overall accuracy of predictions. This approach becomes complex when attempting to identify the precise criteria that make PD or TD model more suitable for a particular region, and these criteria may vary based on model configurations, training durations, or the quality of the available crop data. Therefore, all decisions should be based on a thorough evaluation of the model's spatiotemporal capabilities.

While our study offers valuable insights, we also acknowledge several limitations to ensure a comprehensive understanding of our findings. First, the selection of the prediction model is crucial; if other machine learning or traditional time-series models are employed, our detailed findings may change. To simplify the analysis of in-season predictability, we utilized a uniform national crop calendar. However, this may overlook regional (sub-national) differences in crop cycles, leading to potential discrepancies in the seasonal alignment of EO data.

This may not have a significant impact on Burkina Faso and Somalia, but it may be crucial for other countries. The prediction performances varied between Burkina Faso and Somalia. It's evident that each country, or even regions within a country, might possess distinct economic, environmental, and geographical attributes that influence agricultural productivity. In fact, many studies, including ours, utilize extensive EO data combined with machine learning for large-scale crop forecasting. However, important agro-economic factors, such as market dynamics and fertilizer trends, which play pivotal roles in SSA's crop production (Bonilla-Cedrez et al., 2021), are often overlooked. Burkina Faso is an example of how improvements in agricultural practices and resources can sometimes outweigh the effects of environmental and climatic factors. In addition, the accuracy and consistency of subnational data frequently vary, particularly in the context of SSA.

Potential improvements and new developments of AI and EO data can enhance their applications to agricultural forecasting and modeling (Nakalembe and Kerner, 2023). Specifically, with the proliferation of remote sensing products and the inherent capacity of PD model to assimilate more spatial data, there may be untapped potential in PD versus TD. The challenge lies in identifying the most efficient spatial characteristics to improve modeling temporal variations. Therefore, further research into the relationships between EO data and spatial agroeconomic or agroecological data is necessary. Additionally, these time-invariant agroeconomic and agroeconomic information may exhibit pseudo-time-invariant characteristics over extended periods. In light of this, further analysis is warranted to examine how these characteristics influence model settings and accuracy. Lastly, incorporating seasonal forecast information to forecast crop yields might enhance predictability and provide longer prediction lead times (Harrison et al., 2022; Jin et al., 2022; Shukla et al., 2021)

6. Conclusions

In this study, we evaluated combinations of different modeling and cross-validation approaches used for operational forecasting of subnational crop production and yield predictions. By leveraging a consistent machine learning framework and integrating EO datasets, we explored diverse sampling methods and training durations to predict maize productions and yields in Burkina Faso and Somalia. Our central aims were to discern the value of time-invariant features within the context of a PD model versus a TD model and to evaluate the PD model's potential to improve operational forecasting. Our principal findings are:

- The TD model generally performed well in predicting both crop production and yield, while the PD model offered comparable yield predictions, benefiting from time-invariant features. Time-invariant features, notably livelihood zones, soil properties, and cropland area, proved valuable in capturing spatial characteristics of crop data. These features enhanced modeling of temporal variability, boosting the R-squared to 0.41 and reducing the MAPE to 29 %.
- In Burkina Faso, a short training period sufficed to model both spatial and temporal variability due to the consistent nature of crop data over time. Meanwhile, in Somalia, the PD model demonstrated resilience against crop production outliers. Particularly in major producing regions, its efficacy heightened with prolonged training periods.
- Filling in missing gaps in crop statistics is feasible using the PD model. Challenges persist in predicting entire records for new districts, but clustering based on agro-ecological characteristics may offer a solution.

Our research demonstrates how EO data and machine learning can be combined to improve model accuracies and how spatial and temporal information can be integrated during model development. A tailored approach that considers the unique characteristics of each region can help account for spatial and temporal variability to increase model accuracies.

Declaration of generative AI in scientific writing

During the preparation of this work, DL used Google Bard for grammar verification and readability improvement. This tool served solely to bolster readability, with the assurance from DL that all insights and originality remain the authors'. DL reviewed and edited the content as needed and takes full responsibility for the content of the publication.

CRediT authorship contribution statement

Donghoon Lee: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft. **Frank Davenport:** Conceptualization, Methodology, Writing – review & editing, Project administration. **Shraddhanand Shukla:** Conceptualization, Methodology, Project administration, Writing – review & editing. **Greg Husak:** Writing – review & editing. **Chris Funk:** Writing – review & editing. **James Verdin:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Donghoon Lee reports financial support was provided by United States Geological Survey. Donghoon Lee reports a relationship with US Geological Survey that includes: funding grants. Shraddhanand Shukla reports a relationship with US Geological Survey that includes: funding grants. Frank Davenport reports a relationship with NASA that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We sincerely thank Mike Budde and James Rowland for their invaluable feedback and comments throughout this research. This work is supported by U.S. Geological Survey, United States grant (#G21AC00026) and the NASA, United States SERVIR program (#80NSSC20K0159). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and the U.S. Geological Survey. The authors acknowledge the Center for Scientific Computing at the California Nanosystems Institute at the University of California, Santa Barbara for the availability of high-performance computing resources and support. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2024.110213.

References

Adah, M.I., Kayode, O.O., Victor, I.C., 2017. Cereal productivity in West Africa: a panel data analysis. Am. J. Bus. Econ. Manag. 5, 38.
Allen, R.G., Pereira, L.S., Raes, D, Smith, M (Eds.), 1998. Crop Evapotranspiration: Guidelines For Computing Crop Water Requirements, FAO Irrigation and Drainage Paper. Food and Agriculture Organization of the United Nations, Rome.
Anderson, P., Davenport, F., Baylis, Kathy, Shukla, S., 2024. Using Earth observation products to predict maize prices in Southern Africa. Glob. Food Secur.

Andreadis, K.M., Das, N., Stampoulis, D., Ines, A., Fisher, J.B., Granger, S., Kawata, J., Han, E., Behrangi, A., 2017. The regional hydrologic extremes assessment system: a software framework for hydrologic modeling and data assimilation. PLoS ONE 12, e0176506. https://doi.org/10.1371/journal.pone.0176506.

Anyamba, A., Tucker, C., 2012. Historical perspectives on AVHRR NDVI and vegetation drought monitoring, in: Remote Sensing of Drought: Innovative Monitoring Approaches.

Baltagi, B.H., 2021. Econometric Analysis of Panel Data, Springer Texts in Business and Economics. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-53953-5.

Beauchamp, E.G., Lathwell, D.J., 1967. Effect of changes in root zone temperature on the subsequent growth and development of young corn plants. Agron. J. 59, 189–193. https://doi.org/10.2134/agronj1967.00021962005900020021x.

Bonilla-Cedrez, C., Chamberlin, J., Hijmans, R.J., 2021. Fertilizer and grain prices constrain food production in Sub-Saharan Africa. Nat. Food 2, 766–772. https://doi.org/10.1038/s43016-021-00370-1.

Brady, N.C., Weil, R.R., 2008. The Nature and Properties of Soils. Rev. 14. ed. ed. Pearson/Prentice Hall, Upper Saddle River, NJ.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Butler, E.E., Huybers, P., 2015. Variations in the sensitivity of US maize yield to extreme temperatures by region and growth phase. Environ. Res. Lett. 10, 034009. https://doi.org/10.1088/1748-9326/10/3/034009.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Conradt, T., Gornott, C., Wechsung, F., 2016. Extending and improving regionalized winter wheat and silage maize yield regression models for Germany: enhancing the predictive skill by panel definition through cluster analysis. Agric. For. Meteorol. 216, 68–81. https://doi.org/10.1016/j.agrformet.2015.10.003.

Davenport, F.M., Harrison, L., Shukla, S., Husak, G., Funk, C., McNally, A., 2019. Using out-of-sample yield forecast experiments to evaluate which earth observation products best indicate end of season maize yields. Environ. Res. Lett. 14, 124095. https://doi.org/10.1088/1748-9326/ab5ccd.

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. Environ. Res. Lett. 16, 084050. https://doi.org/10.1088/1748-9326/ac15cc.

Delincé, J., 2017. Recent Practices and Advances for AMIS Crop Yield Forecasting at Farm and Parcel Level: A Review. FAO, Rome, Italy.

FEWS NET, 2023. Peak population in need of emergency food assistance in 2023.

FEWS NET, 2011. Livelihoods zoning "plus" activity in Niger.

Frees, E., 2004. Longitudinal and Panel Data: Analysis and Applications for the Social Sciences. Cambridge University Press.

Fritz, S., See, L., Bayas, J.C.L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J., Rembold, F., Rojas, O., Schucknecht, A., Van Der Velde, M., Verdin, J., Wu, B., Yan, N., You, L., Gilliams, S., Mücher, S., Tetrault, R., Moorthy, I., McCallum, I., 2019. A comparison of global agricultural monitoring systems and current gaps. Agric. Syst. 168, 258–272. https://doi.org/10.1016/j.agsy.2018.05.010.

Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., Abdel Aziz, S., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D.R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., Van Der Velde, M., Dunwoody, A., Kraxner, F., Obersteiner, M., 2015. Mapping global cropland and field size. Glob. Change Biol. 21, 1980–1992. https://doi.org/10.1111/gcb.12838.

Funk, C., Dettinger, M.D., Michaelsen, J.C., Verdin, J.P., Brown, M.E., Barlow, M., Hoell, A., 2008. Warming of the Indian Ocean threatens eastern and southern African food security but could be mitigated by agricultural development. Proc. Natl. Acad. Sci 105, 11081–11086. https://doi.org/10.1073/pnas.0708196105.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. Sci. Data 2, 150066. https://doi.org/10.1038/sdata.2015.66.

Funk, C., Shukla, S., Thiaw, W.M., Rowland, J., Hoell, A., McNally, A., Husak, G., Novella, N., Budde, M., Peters-Lidard, C., Adoum, A., Galu, G., Korecha, D., Magadzire, T., Rodriguez, M., Robjhon, M., Bekele, E., Arsenault, K., Peterson, P., Harrison, L., Fuhrman, S., Davenport, F., Landsfeld, M., Pedreros, D., Jacob, J.P., Reynolds, C., Becker-Reshef, I., Verdin, J., 2019. Recognizing the famine early warning systems network: over 30 years of drought early warning science advances and partnerships promoting global food security. Bull. Am. Meteorol. Soc. 100, 1011–1027. https://doi.org/10.1175/BAMS-D-17-0233.1.

Harrison, L., Landsfeld, M., Husak, G., Davenport, F., Shukla, S., Turner, W., Peterson, P., Funk, C., 2022. Advancing early warning capabilities with CHIRPS-compatible NCEP GEFS precipitation forecasts. Sci. Data 9, 375. https://doi.org/10.1038/s41597-022-01468-2.

Hatfield, J.L., Boote, K.J., Kimball, B.A., Ziska, L.H., Izaurralde, R.C., Ort, D., Thomson, A.M., Wolfe, D., 2011. Climate impacts on agriculture: implications for crop production. Agron. J. 103, 351–370. https://doi.org/10.2134/agronj2010.0303.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km — global soil information based on automated mapping. PLoS ONE 9, e105992. https://doi.org/10.1371/journal.pone.0105992.

Hillel, D.I., 1982. Introduction to Soil Physics. Academic Press, New York.

Hobbins, M., 2016. The variability of ASCE standardized reference evapotranspiration: a rigorous, CONUS-Wide decomposition and attribution. Trans. ASABE 59, 561–576. https://doi.org/10.13031/trans.59.10975.

Hsiao, C., 2007. Panel data analysis—Advantages and challenges. TEST 16, 1–22. https://doi.org/10.1007/s11749-007-0046-x.

Hu, B., McAleer, M., 2005. Estimation of Chinese agricultural production efficiencies with panel data. Math. Comput. Simul. 68, 474–483. https://doi.org/10.1016/j.matcom.2005.02.002.

Jensen, M.E., Burman, R.D., Allen, R.G., American Society of Civil Engineers, 1990. Evapotranspiration and irrigation water requirements: a manual. ASCE Manuals and Reports on Engineering Practice. The Society, New York, N.Y.

Jin, H., Li, M., Hopwood, G., Hochman, Z., Bakar, K.S., 2022. Improving early-season wheat yield forecasts driven by probabilistic seasonal climate forecasts. Agric. For. Meteorol. 315, 108832. https://doi.org/10.1016/j.agrformet.2022.108832.

Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. Science 304, 1623–1627. https://doi.org/10.1126/science.1097396.

Lala, J., Yang, M., Wang, G., Block, P., 2021. Utilizing rainy season onset predictions to enhance maize yields in Ethiopia. Environ. Res. Lett. 16, 054035. https://doi.org/10.1088/1748-9326/abf9c9.

Laudien, R., Schauberger, B., Makowski, D., Gornott, C., 2020. Robustly forecasting maize yields in Tanzania based on climatic predictors. Sci. Rep. 10, 19650. https://doi.org/10.1038/s41598-020-76315-8.

Laudien, R., Schauberger, B., Waid, J., Gornott, C., 2022. A forecast of staple crop production in Burkina Faso to enable early warnings of shortages in domestic food availability. Sci. Rep. 12, 1638. https://doi.org/10.1038/s41598-022-05561-9.

Lee, D., Davenport, F., Shukla, S., Husak, G., Funk, C., Harrison, L., McNally, A., Rowland, J., Budde, M., Verdin, J., 2022. Maize yield forecasts for Sub-Saharan Africa using Earth observation data and machine learning. Glob. Food Secur. 33, 100643. https://doi.org/10.1016/j.gfs.2022.100643.

Lee, D., Anderson, W., Chen, X., Davenport, F., Shukla, S., Sahajpal, R., Budde, M., Rowland, J., Verdin, J., You, L., Ahouangbenon, M., Davis, K.F., Kebede, E., Ehrmann, S., Justice, C., Meyer, C., 2024. HarvestStat Africa – harmonized subnational crop statistics for Sub-Saharan Africa. EarthArXiv. https://doi.org/10.31223/X5M123.

Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate trends and global crop production since 1980. Science 333, 616–620. https://doi.org/10.1126/science.1204531.

Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. Remote Sens. Environ. 164, 324–333. https://doi.org/10.1016/j.rse.2015.04.021.

Lu, J., Carbone, G.J., Huang, X., Lackstrom, K., Gao, P., 2020. Mapping the sensitivity of agriculture to drought and estimating the effect of irrigation in the United States, 1950–2016. Agric. For. Meteorol. 108124, 292–293. https://doi.org/10.1016/j.agrformet.2020.108124.

Maman, G., Idriss, S., Wortmann, C., 2018. Crop yield response to fertilizer relative to soil properties in Sub-Saharan Africa. Soil Sci. Soc. Am. J. 82, 862–870. https://doi.org/10.2136/sssaj2018.02.0066.

Mcmaster, G., 1997. Growing degree-days: one equation, two interpretations. Agric. For. Meteorol. 87, 291–300. https://doi.org/10.1016/S0168-1923(97)00027-0.

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C.D., Verdin, J.P., 2017. A land data assimilation system for sub-Saharan Africa food and water security applications. Sci. Data 4, 170012. https://doi.org/10.1038/sdata.2017.12.

McNally, A., McCartney, S., Ruane, A.C., Mladenova, I.E., Whitcraft, A.K., Becker-Reshef, I., Bolten, J.D., Peters-Lidard, C.D., Rosenzweig, C., Uz, S.S., 2019. Hydrologic and agricultural earth observations and modeling for the water-food nexus. Front. Environ. Sci. 7, 23. https://doi.org/10.3389/fenvs.2019.00023.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Softw. 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001.

Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. 391, 202–216. https://doi.org/10.1016/j.jhydrol.2010.07.012.

Mitra, P., Unsal, F., Farid, M., Kemoe, L., Fayad, D., Spray, J., Okou, C., Salgado Baptista, D.M., Lanci, L., Muehlschlegel, T., Tuitoek, K., 2022. Climate change and chronic food insecurity in Sub-Saharan Africa. Dep. Pap. 2022, 1. https://doi.org/10.5089/9798400218507.087.

Monfreda, C., Ramankutty, N., Foley, J.A., 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Glob. Biogeochem. Cycles 22, 2007GB002947. https://doi.org/10.1029/2007GB002947.

Nakalembe, C., Becker-Reshef, I., Bonifacio, R., Hu, G., Humber, M.L., Justice, C.J., Keniston, J., Mwangi, K., Rembold, F., Shukla, S., Urbano, F., Whitcraft, A.K., Li, Y., Zappacosta, M., Jarvis, I., Sanchez, A., 2021. A review of satellite-based global agricultural monitoring systems available for Africa. Glob. Food Secur. 29, 100543. https://doi.org/10.1016/j.gfs.2021.100543.

Nakalembe, C., Kerner, H., 2023. Considerations for AI-EO for agriculture in Sub-Saharan Africa. Environ. Res. Lett. 18, 041002. https://doi.org/10.1088/1748-9326/acc476.

Porter, J.R., Gawith, M., 1999. Temperatures and the growth and development of wheat: a review. Eur. J. Agron. 10, 23–36. https://doi.org/10.1016/S1161-0301(98)00047-1.

Rawls, W.J., Brakensiek, D.L., Saxtonn, K.E., 1982. Estimation of soil water properties. Trans. ASAE 25, 1316–1320. https://doi.org/10.13031/2013.33720.

Schauberger, B., Jägermeyr, J., Gornott, C., 2020. A systematic review of local to regional yield forecasting approaches and frequently used data resources. Eur. J. Agron. 120, 126153. https://doi.org/10.1016/j.eja.2020.126153.

Shukla, S., Landsfeld, M., Anthony, M., Budde, M., Husak, G.J., Rowland, J., Funk, C., 2021. Enhancing the application of earth observations for improved environmental decision-making using the early warning explorer (EWX). Front. Clim. 2, 583509. https://doi.org/10.3389/fclim.2020.583509.

Skakun, S., Justice, C.O., Vermote, E., Roger, J.-C., 2018. Transitioning from MODIS to VIIRS: an analysis of inter-consistency of NDVI data sets for agricultural monitoring. Int. J. Remote Sens. 39, 971–992. https://doi.org/10.1080/01431161.2017.1395970.

Turner, W.A., Husak, G., Funk, C., Roberts, D.A., Jones, C., 2022. An improved climatological forecast method for projecting end-of-season water requirement satisfaction index. J. Hydrometeorol. 23, 1281–1295. https://doi.org/10.1175/JHM-D-21-0184.1.

Van Klompenburg, T., Kassahun, A., Catal, C., 2020. Crop yield prediction using machine learning: a systematic literature review. Comput. Electron. Agric. 177, 105709. https://doi.org/10.1016/j.compag.2020.105709.

Verdin, J., Funk, C., Senay, G., Choularton, R., 2005. Climate science and famine early warning. Philos. Trans. R. Soc. B Biol. Sci. 360, 2155–2168. https://doi.org/10.1098/rstb.2005.1754.

Wang, Y., Khodadadzadeh, M., Zurita-Milla, R., 2023. Spatial+: a new cross-validation method to evaluate geospatial machine learning models. Int. J. Appl. Earth Obs. Geoinf. 121, 103364. https://doi.org/10.1016/j.jag.2023.103364.

Wooldridge, J., 2010. Econometric Analysis of Cross Section and Panel Data, 2nd ed. The MIT Press.

You, L., Rosegrant, M.W., Wood, S., Sun, D., 2009. Impact of growing season temperature on wheat productivity in China. Agric. For. Meteorol. 149, 1009–1014. https://doi.org/10.1016/j.agrformet.2008.12.004.