remote sensing

MDPI

Technical Note

Mapping Field-Level Maize Yields in Ethiopian Smallholder Systems Using Sentinel-2 Imagery

Zachary Mondschein [1], Ambica Paliwal [1,2], Tesfaye Shiferaw Sida [3], Jordan Chamberlin [4], Runzi Wang [1] and Meha Jain [1,*]

[1] School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48109, USA; zrmondsc@umich.edu (Z.M.); ambica@umich.edu (A.P.); runziw@umich.edu (R.W.)
[2] International Livestock Research Institute (ILRI), Nairobi 00100, Kenya
[3] International Maize and Wheat Improvement Center (CIMMYT), Addis Ababa P.O. Box 5689, Ethiopia; t.sida@cgiar.org
[4] International Maize and Wheat Improvement Center (CIMMYT), Nairobi 00621, Kenya; j.chamberlin@cgiar.org
* Correspondence: mehajain@umich.edu

Abstract: Remote sensing offers a low-cost method for estimating yields at large spatio-temporal scales. Here, we examined the ability of Sentinel-2 satellite imagery to map field-level maize yields across smallholder farms in two regions in Oromia district, Ethiopia. We evaluated how effectively different indices, the MTCI, GCVI, and NDVI, and different models, linear regression and random forest regression, can be used to map field-level yields. We also examined if models improved by adding weather and soil data and how generalizable our models were if trained in one region and applied to another region, where no data were used for model calibration. We found that random forest regression models that used monthly MTCI composites led to the highest yield prediction accuracies (R^2 up to 0.63), particularly when using only localized data for training the model. These models were not very generalizable, especially when applied to regions that had significant haze remaining in the imagery. We also found that adding soil and weather data did little to improve model fit. Our results highlight the ability of Sentinel-2 imagery to map field-level yields in smallholder systems, though accuracies are limited in regions with high cloud cover and haze.

Keywords: Sentinel-2; yield mapping; smallholder farms; agriculture; maize

Citation: Mondschein, Z.; Paliwal, A.; Sida, T.S.; Chamberlin, J.; Wang, R.; Jain, M. Mapping Field-Level Maize Yields in Ethiopian Smallholder Systems Using Sentinel-2 Imagery. *Remote Sens.* **2024**, *16*, 3451. https://doi.org/10.3390/rs16183451

Academic Editors: Jochem Verrelst, Katja Berger, Egor Prikaziuk and Clement Atzberger

Received: 14 June 2024
Revised: 30 August 2024
Accepted: 31 August 2024
Published: 18 September 2024

1. Introduction

Efforts to meet global food demand in the coming decades will be challenged by population growth and climate change [1,2]. One way to meet this growing demand is to increase agricultural production. Yet increasing agricultural production through agricultural extensification is associated with various environmental costs, such as greenhouse gas emissions and biodiversity loss [1]. Instead, closing yield gaps, or narrowing the gap between current agricultural yields and potential agricultural yields on existing land, could be a way to meet future food demand more sustainably [3]. This is especially important in regions such as sub-Saharan Africa (SSA), where yield gaps are large [4], climatic impacts are severe [5], and population growth is rapid [6]. In particular, Ethiopia, which is the second most populous country in SSA, is one of the countries in SSA that is most food insecure and vulnerable to climate change [7,8]. Closing yield gaps for maize will be especially important in this region, given that maize provides nearly 20% of the nation's calories [9] and is one crop that is projected to be the most negatively impacted by climate change [6].

In order to identify the causes of and potential solutions to close yield gaps, we must be able to reliably estimate yields across large spatial and temporal scales [3]. Yet, to date, this has been challenging through on-the-ground data collection efforts. This is because such

on-the-ground surveys are expensive, difficult to conduct at larger spatial scales, and tend to rely on often inaccurate, self-reported data [10,11]. A potential low-cost way to produce agricultural statistics at scale, and gain an understanding of yield gaps, is to use remote sensing [12]. However, mapping field-level yields in smallholder systems such as Ethiopia can be challenging given small field sizes (<1 ha) and high within-field variability and between-field variability due to heterogeneity of management practices and environmental conditions. The launch of high spatial and temporal resolution satellites, such as the Sentinel constellations and PlanetScope, have helped overcome such challenges [13,14]. For example, several recent studies have demonstrated the potential of optical Sentinel-2 data to map field-level yields in heterogeneous smallholder systems [15–17]. Sentinel-2 imagery also has frequent revisit times (5/day), which have been linked with higher accuracy when mapping yields [13,18].

Various vegetation indices have been used to map crop yields, and efficacy may depend on crop type and local conditions [19]. While the Normalized Difference Vegetation Index (NDVI) has been used extensively to map yields [19–21], vegetation indices that use the green rather than the red band to optimize for chlorophyll sensitivity may be more reliable for crop yield estimation [22,23]. Specifically, previous studies have shown that the Green Chlorophyll Vegetation Index (GCVI), which is more sensitive to moderate to high levels of canopy chlorophyll, can outperform traditional vegetation indices such as the NDVI [19,24,25]. In addition, the Medium Resolution Imaging Spectrophotometer (MERIS) Terrestrial Chlorophyll Index (MTCI), which uses red-edge reflectance, is sensitive to canopy chlorophyll and nitrogen content and has also been shown to outperform the NDVI and GCVI in previous comparisons [22,24,26]. However, MTCI indices generated from Sentinel-2 imagery have a coarser spatial resolution than the NDVI and GCVI, owing to the reliance on the red edge (RE) band (band 5), which is provided at 20 m instead of 10 m spatial resolution. Thus, the potential of the MTCI to map crop yields at the field level in small fields should be further investigated.

Previous studies have used ground-based yield measurements, such as crop cuts, to train linear regression models that translate vegetation indices into yield estimates [13,25]. Linear regression models have been favored for their simplicity and ease of implementation, especially when applied in cloud computing platforms such as Google Earth Engine (GEE), and because of the observed linear relationships between vegetation indices and yield [27,28]. However, it is possible that machine learning models, such as random forest, may outperform such simple linear regression models as they can account for complex interactions among explanatory variables [29]. They are also considered to be resistant to overfitting [30].

There is also concern that models that require ground calibration are limited in their scalability [31]. This is because the relationship between vegetation indices and yields varies depending on region, crop variety, and management practices [27]. Thus, it is suggested that these models should be recalibrated using new ground data before they are applied to other geographic areas [26]. It is unclear to what extent such models can be extrapolated to estimate spatial patterns in yield outside the region in which they are calibrated. Approaches that utilize crop growth models have been advanced as a possible solution [27]. However, these approaches are computationally intensive, and require various agro-meteorological inputs [32]. Satellite-based crop yield models that rely on ground data should thus be validated on data outside of their geographic area before extrapolating over larger regions.

Besides vegetation indices, climate data have long been used to understand crop yields. Climate variables, like air temperature and precipitation, have been shown to explain up to one-third of the variation in country-level crop production [33]. Leroux et al. (2016) suggest that vegetation indices alone cannot account for differences in the relationship between aboveground biomass and harvestable yield, which may be caused by environmental stressors [34]. Satellite-based soil moisture and evapotranspiration measurements have also been found to be more closely correlated to yield than vegetation index measurements,

and better at quantifying water stress [35]. To address these limitations, several studies have examined machine learning yield estimation using a mix of vegetation indices and remote sensing derived climate variables or soil variables [34,36–39]. Many of these studies are concerned with yield estimates at higher levels of spatial aggregation (e.g., country, state/province, or county level). However, less research has focused on how climate data might augment the accuracy and generalizability of vegetation index-based yield estimation at fine spatial scales, such as the field level. This is likely because many climate datasets are only available at relatively coarse spatial resolutions, often larger than individual field sizes in smallholder systems.

This study adds to the growing body of work that assesses the ability of using high spatio-temporal resolution satellite data to map field-level yields in smallholder systems at scale. Our study is one of the first to map field-level maize yields in Ethiopia using Sentinel-2 imagery. Previous work in Ethiopia has typically used coarse-resolution sensors (e.g., MODIS) to map maize yield (e.g., Debalke and Abebe, 2022) at scales coarser than the field or have used indices that do not rely on the red-edge band (e.g., Guo et al., 2023), which have been shown to be superior for yield estimation in other studies [40,41]. We compare multiple methods to estimate maize yield in two regions within Oromia district, Ethiopia, during the 2021 growing season. In this study, we specifically examine:

1. How well can we map field-level yields in smallholder maize systems in Ethiopia using Sentinel-2 imagery?
2. Which vegetation index results in the highest yield prediction accuracies: the NDVI, GCVI, or MTCI?
3. Which model leads to higher prediction accuracies: multiple linear regression or random forest regression?
4. Does imputing missing values due to cloud cover improve model performance?
5. Can adding weather and soil data improve prediction accuracy compared to using only vegetation indices?
6. Is it possible to create a generalizable model that accurately estimates yields across multiple regions using limited ground data for training?

Our results provide important insights into the ability of Sentinel-2 imagery to map field-level yields in heterogeneous smallholder systems. This is critically important as smallholder systems are projected to face some of the largest increases in food demand over the coming decades [6], and such yield information can help identify where yield gaps are the largest and potential interventions that may help close these yield gaps.

2. Study Area

Our study area spans a 30,000 km^2 region in Oromia district, Ethiopia (Figure 1), with data collected in two distinct sub-regions. The first sub-region comprises an approximately 8800 km^2 area straddling parts of the East Shewa and Guraghe Zones (Figure 1b). The second sub-region comprises a 1700 km^2 area in the western Jimma Zone (Figure 1c). The greater region is dominated by smallholder agriculture, with cropland covering over 72% of the land area. There are two agricultural growing seasons in Ethiopia when maize is grown: the long rainy season (Meher) and the short rainy season (Belg) [42]. The focus of this study is the long rainy season of Meher, when the majority of maize is grown, which spans from April to September [43].

Agricultural management practices and soil conditions are highly heterogeneous over the study region. Specifically, 81% of surveyed fields were less than 5 hectares, and approximately one-quarter of all fields (26%) were less than two hectares. In addition, 93% of farmers applied fertilizer, such as DAP and urea, but inputs varied from 0 to 300 kg throughout the season. There was likewise a diversity of maize varieties planted, with the most common varieties being BH661, Limu, Damote, and Shone. Soil texture was silt, clay, or sand, with silt being the most common (40%). A majority of fields (67%) had fruiting trees, but only 20% of fields were intercropped with other species, of which beans were the most common. None of the fields surveyed used irrigation. Although sow date and

harvest date varied slightly, most fields (90%) were sown in April or May, and all fields were harvested in November.

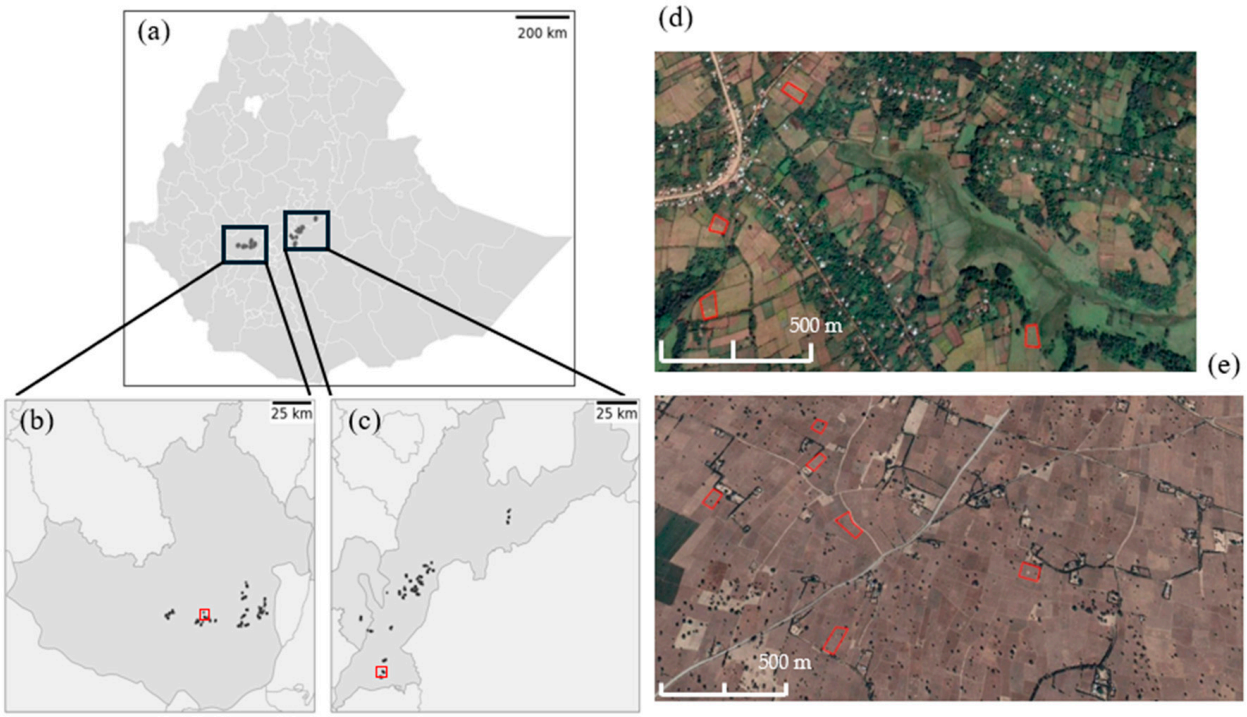

Figure 1. Map of the study area with (**a**) a countrywide map of Ethiopia with the field locations, (**b**) Jimma zone with field locations, (**c**) East Shewa and Guraghe zones with field locations, (**d**) details of the red rectangle in panel b showing field boundaries over high-resolution aerial imagery, and (**e**) details of the red rectangle in panel (**c**) showing field boundaries in red over high-resolution aerial imagery. Aerial imagery via Google Earth 7.3.6 (2023) CNES/Airbus [Accessed on 18 April 2023].

3. Methods

We processed Sentinel-2 imagery (Section 3.2) to estimate maize yields across fields where we collected on-the-ground crop cut yield information (Question 1, Section 3.1). To answer our main research questions, we calculated three different vegetation indices (Question 2) from Sentinel-2 satellite images at two different levels of temporal aggregation throughout the 2021 maize growing season for inclusion in maize yield models (Section 3.2). For each vegetation index, we also compared two different yield estimation models, linear regression and random forest regression (Question 3, Section 3.4). We also examined whether imputing missing values due to cloud cover led to improved model performance (Question 4, Section 3.2). To understand if adding weather and soil data improved model fit, we extracted field-level statistics for five different environmental variables from gridded climate and soil data products (Section 3.3). We recalibrated each model using an expanded feature space that included both vegetation index and environmental data and compared the performance of this enhanced model to the baseline model with no environmental data (Question 5, Section 3.4). We examined how generalizable our yield estimation algorithms are across the full study region by training a model using data from only one region and then applying and validating it in the other region (Question 6, Section 3.5).

3.1. Crop Cut Data

Most methods for estimating crop yield use field-level yield data to calibrate remote sensing models [13]. The gold standard for yield estimation in field is through collecting crop cuts, which we used to estimate yield at the end of the growing season [44]. Agricultural surveys and crop cuts were administered by collaborators working with the International Maize and Wheat Improvement Centre for 406 fields across our study area

in 2021. Each maize field was split into four quadrants. At the center of each quadrant, a 5 m × 8 m area was harvested. The maize was dried and threshed before it was weighed in field. We averaged all subplot yields to obtain field-level yields. It is important to note that even though above-ground dry matter (e.g., biomass) would likely be better correlated with satellite vegetation indices, we focused on predicting yield as this is the measure that farmers and agricultural development organizations prioritize for estimating overall production and yield gaps. Differences in harvest index due to different varieties and management likely exist and will be captured as residual errors in our models. GPS coordinates were collected for each crop cut sub-plot, as well as for the four corners of the field boundaries. Field boundary polygons were constructed in Python using the Shapely package version 2.0.1 [45]. The boundaries were then manually corrected in Google Earth Pro by aligning field boundaries with visible boundaries from the latest available high-resolution aerial imagery in Google Earth Pro. Field boundaries were described as 'low', 'medium', or 'high' confidence based on how closely they corresponded to visible boundaries in the aerial imagery. We retained only the 321 'high' confidence fields for our analysis, comprising 158 fields in the Jimma sub-region and 163 fields in the East Shewa-Garaghe sub-region.

3.2. Sentinel-2 Imagery

We accessed Level 2A, atmospherically corrected surface reflectance Sentinel-2 imagery through the Google Earth Engine (GEE) platform [46]. As the growing season coincided with the rainy season, many of the Sentinel-2 images had a high number of cloudy pixels. Pixel-wise cloud masking was performed in GEE using the cloud probability score generated by the Sentinel Hub Cloud Detector algorithm using the s2Cloudless Python library [47]. The Sentinel Hub Cloud Detector is a readily available machine learning cloud and cloud shadow masking algorithm for use in conjunction with Sentinel-2 surface reflectance imagery. Based on visual inspection of cloud removal, we set the cloud probability threshold parameter at 40% and the cloud filter threshold parameter to 100%. As a result, every image from the Sentinel-2 surface reflectance image collection from 15 March 2021 to 5 December 2021 was retained for our analysis, regardless of cloud cover percentage, and pixels with a cloud probability score greater than 40% were masked.

Three vegetation indices commonly used in satellite-based yield estimation studies, namely the NDVI, GCVI, and MTCI, were calculated for each Sentinel-2 image following the application of the cloud mask. The relevant Sentinel-2 bands used to calculate each vegetation index are recorded in Table 1, and the formula for each index is listed in Table 2. We then created maximum vegetation index composites at two different levels of temporal aggregation using the 'qualityMosaic' function in GEE. We created 15-day 'biweekly' and 30-day 'monthly' image composites across the entire growing season for each vegetation index. We then applied the 'reduceRegions' method to obtain field-level mean VI for each composite image. In cases where more than 80% of pixels were removed from a given field due to cloud cover, we set the value to null and instead imputed the value using Equation (1):

$$VI_t = VI_{t-x} + (VI_{t+y} - VI_{t-x})/(x + y) \quad (1)$$

where VI_t is a null observation at time (t), VI_{t-x} is the first non-null observation that occurred prior to VI_t in the time series at x days before time t, VI_{t+y} is the first non-null observation that occurred after VI_t in the time series at y days after time t, and x + y represents the sum of the number of days that occurred before and after time t.

Table 1. Sentinel-2 bands used for the computation of vegetation indices.

Band	Spectral Range (nm)	Resolution (m)
Green	543–578	10
Red	650–680	10
Red Edge (RE)	690–730	20
Near Infrared (NIR)	760–850	10

Table 2. Three vegetation indices used in this study and their formulae [48–50].

Vegetation Index	Formula	Reference
Normalized Difference Vegetation Index (NDVI)	(NIR − Red)/(NIR + Red)	Rouse et al., 1973 [48]
Green Chlorophyll Vegetation Index (GCVI)	(NIR/Green − 1)	Gitelson et al., 2003 [49]
MERIS Terrestrial Chlorophyll Index (MTCI)	(NIR − RE)/(RE − Red)	Dash & Curran, 2004 [50]

The above steps resulted in six different sets of predictor variables for each sub-region, with each set comprising a time series of one of three vegetation indices (the NDVI, GCVI, or MTCI) at a given temporal aggregation level (biweekly or monthly). The three biweekly VI predictor sets for the Jimma sub-region and the full Oromia region were discarded as there was too much cloud cover to extract data for a majority of fields. To quantify the impact of imputing missing values on model performance, we also removed fields that had missing values in the East Shewa–Garaghe sub-region and reran our analyses with no imputation. We only conducted this comparison analysis for the Eastern region as it had only a few observations (n = 15 field/months) that had missing values at the monthly level.

3.3. Environmental Data

We included five gridded environmental data products: (1) maximum Land Surface Temperature (LST_{max})—calculated from the daily 1 km resolution Daytime LST thermal product from the Terra Moderate Resolution Imaging Spectroradiometer (MODIS) sensor (MOD11A1) [51], (2) mean Land Surface Temperature (LST_{mean})—calculated as the mean of the daily 1 km resolution Daytime LST and Nighttime LST product from the Terra MODIS sensor, (3) total precipitation—extracted from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) 0.05° × 0.05° resolution daily precipitation product [52], (4) soil organic carbon—derived from the SoilGrids250m resolution machine learning based product at 0–5 cm depth [53], and (5) total nitrogen—also derived from SoilGrids250m at 0–5 cm depth. All environmental data products were accessed and processed using Google Earth Engine, and field-level statistics were calculated for the period between 15 March 2021 and 15 December 2021 using the 'reduceRegions' method in GEE.

3.4. Model Parameterization and Validation

We compared two classes of models to estimate yield, a linear regression model and a random forest regression model. The linear regression model was estimated using Equation (2):

$$Yield = \beta_0 + \beta_1 VI_1 + \beta_2 VI_2 + \ldots + \beta_n VI_n + \epsilon \quad (2)$$

where *Yield* is the observed yield (kg/ha) at the field scale calculated via crop cuts, $\beta_1 VI_1$ represents the coefficient for the mean vegetation index (*VI*) value for each field for the first composite window, and $\beta_n VI_n$ represents the coefficient for the mean VI value for each field for the n^{th} composite window in the predictor set. The linear regressions were performed in Python using the 'sklearn.linear_model.LinearRegression' routine from the Scikit-Learn

version 1.2.2 machine learning library [54]. We trained individual linear regression models for each of the VI datasets (Section 3.2) with and without the inclusion of environmental data (Section 3.3).

For each linear regression model, we also ran a corresponding random forest regression model using the same set of predictor variables. Random forest is an ensemble learning method that creates multiple decision trees using randomly drawn subsamples of the data [55]. The average is calculated across all decision trees to output a final model. We developed the random forest regression models using the 'sklearn.ensemble.RandomForestRegressor' routine from the Scikit-learn Python module. For each model, we tuned four hyperparameters: the number of trees (num_estimators), the number of features (max_features), the maximum depth of a given tree (max_depth), and the minimum number of samples to split an internal node (min_sample_split). We defined a dictionary with several discrete values for each hyperparameter and used a randomized search with 5-fold cross-validation and 500 iterations to determine optimal hyperparameters based on model R^2 scores [56].

All models were validated using a 70:30 train–test split. The models were scored on the basis of their coefficient of determination (R^2) and root mean squared error (RMSE) when compared to the observed crop cut yield estimates for each field.

3.5. Comparison of Models by Sub-Region

We assessed the spatial generalizability of our modeling approaches by validating the model developed in one sub-region on observations from the other sub-region. Generalizability was assessed only for monthly VI-composite models due to the lack of biweekly VI-composite models in the Jimma sub-region. For simplicity, we also focused only on using the best-performing VI, which was the MTCI, for our generalizability analysis. Accuracy was assessed using R^2 and RMSE values.

4. Results

With respect to our main research question of how well we can predict field-level yields in smallholder maize systems in Ethiopia using Sentinel-2 imagery (Question 1), we were able to map field-level yields well with R^2 ranging from 0.23 (Table 3) to 0.63 (Table 4) across all models. The accuracy of our models varied greatly by region. We observed much higher accuracies in the East Shewa–Guraghe sub-region, where the best-performing model had an R^2 value of 0.63 and an RMSE equal to 1329 kg/ha. In the Jimma sub-region, however, the best-performing model had an R^2 of only 0.35 and an RMSE value of 1746 kg/ha (Figure 2).

Table 3. R^2 and RMSE values for each model vary by region, regression model, vegetation index, and temporal resolution.

Region/ Sub-Region	Regressor	Vegetation Index (VI)	Temporal Resolution	Coefficient of Determination (R^2)	Root Mean Squared Error (RMSE)
Regional	Linear	GCVI	Monthly	0.11	2057
Regional	Linear	MTCI	Monthly	0.23	1913
Regional	Linear	NDVI	Monthly	0.17	1988
Regional	Random Forest	GCVI	Monthly	0.16	1997
Regional	Random Forest	MTCI	Monthly	0.28	1846

Table 3. *Cont.*

Region/ Sub-Region	Regressor	Vegetation Index (VI)	Temporal Resolution	Coefficient of Determination (R^2)	Root Mean Squared Error (RMSE)
Regional	Random Forest	NDVI	Monthly	0.09	2077
East Shewa-Guraghe	Linear	GCVI	Biweekly	0.35	1788
East Shewa-Guraghe	Linear	MTCI	Biweekly	0.49	1579
East Shewa-Guraghe	Linear	NDVI	Biweekly	0.34	1806
East Shewa-Guraghe	Random Forest	GCVI	Biweekly	0.43	1668
East Shewa-Guraghe	Random Forest	MTCI	Biweekly	0.5	1574
East Shewa-Guraghe	Random Forest	NDVI	Biweekly	0.35	1788
East Shewa-Guraghe	Linear	GCVI	Monthly	0.45	1647
East Shewa-Guraghe	Linear	MTCI	Monthly	0.47	1621
East Shewa-Guraghe	Linear	NDVI	Monthly	0.4	1716
East Shewa-Guraghe	Random Forest	GCVI	Monthly	0.4	1713
East Shewa-Guraghe	Random Forest	MTCI	Monthly	0.49	1576
East Shewa-Guraghe	Random Forest	NDVI	Monthly	0.22	1955
Jimma	Linear	GCVI	Monthly	0.17	1969
Jimma	Linear	MTCI	Monthly	0.32	1782
Jimma	Linear	NDVI	Monthly	0.17	1974
Jimma	Random Forest	GCVI	Monthly	0.19	1946
Jimma	Random Forest	MTCI	Monthly	0.35	1746
Jimma	Random Forest	NDVI	Monthly	0.02	2149

Table 4. R^2 and RMSE values for yield models in the East Shewa–Guraghe sub-regions with no imputation using monthly VI composites.

Region/ Sub-Region	Regressor	Vegetation Index (VI)	Temporal Resolution	Coefficient of Determination (R^2)	Root Mean Squared Error (RMSE)
East Shewa-Guraghe	Linear	GCVI	Monthly	0.41	1663
East Shewa-Guraghe	Linear	MTCI	Monthly	0.57	1422
East Shewa-Guraghe	Linear	NDVI	Monthly	0.33	1776
East Shewa-Guraghe	Random Forest	GCVI	Monthly	0.36	1732
East Shewa-Guraghe	Random Forest	MTCI	Monthly	0.63	1326
East Shewa-Guraghe	Random Forest	NDVI	Monthly	0.26	1875

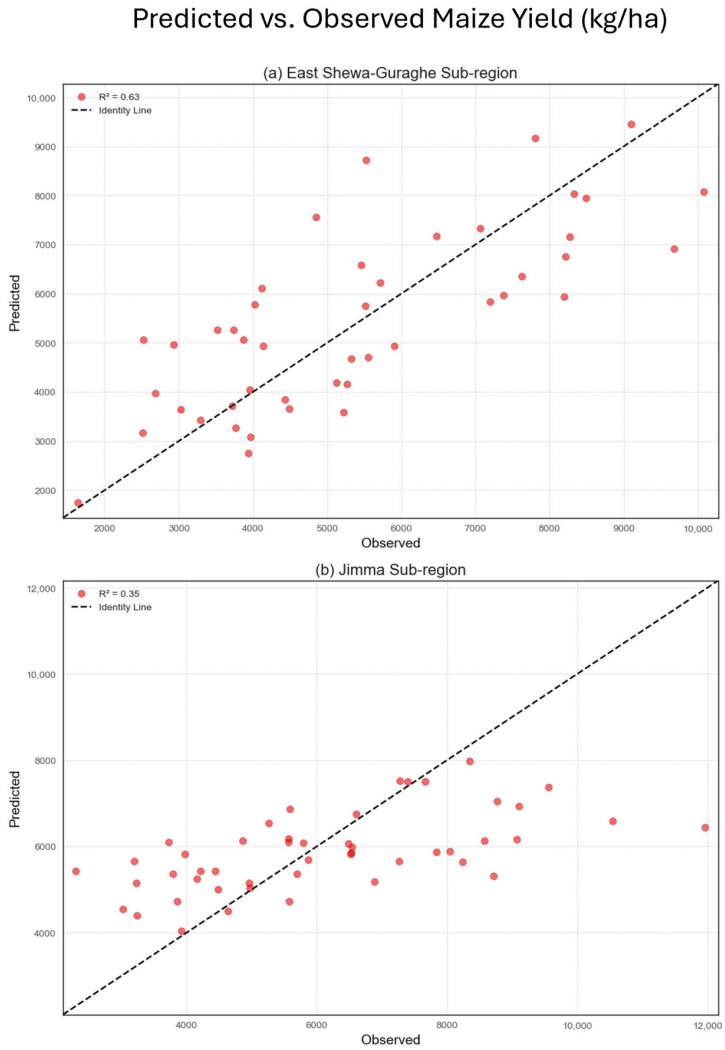

Figure 2. Scatterplots comparing model-estimated maize yield with the yield derived from crop cuts for the best-performing models in (**a**) the East Shewa–Guraghe sub-region and (**b**) the Jimma sub-region. The dashed line represents the 1:1 identity line.

Comparing the three different indices used (Question 2), the GCVI, NDVI, and MTCI, we found that overall, the MTCI led to the highest prediction accuracies across all models and locations, as evidenced by higher R^2 values and lower RMSEs. After the MTCI, the GCVI performed best in random forest models, and the GCVI and NDVI performed similarly in linear regression models. The best-performing model at the regional scale was the random forest regression model with monthly MTCI composite predictors (R^2 = 0.28, RMSE = 1846 kg/ha). The best-performing model in the East Shewa–Guraghe sub-region was the random forest regression model with biweekly MTCI composite predictors (R^2 = 0.50, RMSE = 1574 kg/ha). The best-performing model in the Jimma sub-region was the random forest regression model with monthly MTCI composite predictors (R^2 = 0.35, RMSE = 1746 kg/ha). Across the random forest regression models, the MTCI improved R^2 values by 0.12 and reduced the RMSE by 163 kg/ha compared to the GCVI, the next best-performing vegetation index.

We next examined whether linear regression models or random forest regression models led to the highest yield prediction accuracies (Question 3). We found that across all locations, random forest models outperformed linear regression models, though the difference in R^2 and RMSE was generally small (R^2 = 0.03 and RMSE = 38 kg/ha). The difference was larger in the East Shewa–Guraghe sub-region where random forest models improved the R^2 by 0.06 and decreased the RMSE by 87 kg/ha on average. The difference was less pronounced in the Jimma sub-region where random forest models improved the R^2 by 0.04 and decreased the RMSE by 38 kg/ha on average. At the regional scale, random forest regression narrowly outperformed linear regression, with the R^2 improving by 0.01 and the RMSE decreasing by 12 kg/ha. Across models with MTCI composite predictor variables, which were the highest-performing models in each region, random forest regression models offered a noticeable increase in performance over linear regression models, improving the R^2 by 0.04 and decreasing the RMSE by 49 kg/ha on average.

We assessed how data availability impacts model performance by comparing models in the East Shewa–Guraghe sub-region where we implemented a weighted moving average imputation technique to models where observations with null field-month values were dropped from the dataset (Question 4, Table 4). For the best-performing vegetation index, the MTCI, regression models that dropped null values improved markedly over those that imputed null values. For the linear regression model, the R^2 increased from 0.47 to 0.53 and the RMSE decreased from 1621 kg/ha to 1488 kg/ha. For the random forest regression model, the R^2 increased from 0.50 to 0.63 and the RMSE decreased from 1576 kg/ha to 1329 kg/ha.

We assessed the difference in performance between models that included environmental variables and the baseline models that did not include environmental variables (Question 5). We found that the models including environmental variables (Table 5) led to negligible differences in performance when compared to our baseline models (Table 3). Overall, the RMSE was 2 kg/ha lower on average in the enhanced linear regression models compared to the baseline linear regression models. The RMSE was 16 kg/ha (1%) lower on average in the enhanced random forest regression models compared to the baseline random forest regression models.

We tested generalizability by validating the models produced in one region on test data from the other region (Question 6). Overall, we found that our models were not very generalizable, particularly when applying the models developed in the East Shewa–Guraghe sub-region to the Jimma sub-region (Table 6). The Jimma sub-region models, however, were more generalizable to the East Shewa–Guraghe sub-region. The best-performing model was the linear regression model, reaching R^2 values of 0.36 (RMSE = 1736 kg/ha). Our results suggest that the random forest regression models were less generalizable than the linear regression models.

Table 5. R^2 and RMSE values for regression models that include environmental variables as predictors.

Region/ Sub-Region	Regressor	Vegetation Index (VI)	Temporal Resolution	Coefficient of Determination (R^2)	Root Mean Squared Error (RMSE)
Regional	Linear	GCVI	Monthly	0.11	2061
Regional	Linear	MTCI	Monthly	0.25	1894
Regional	Linear	NDVI	Monthly	0.14	2021
Regional	Random Forest	GCVI	Monthly	0.19	1966
Regional	Random Forest	MTCI	Monthly	0.33	1786
Regional	Random Forest	NDVI	Monthly	0.11	2054
East Shewa-Guraghe	Linear	GCVI	Biweekly	0.43	1672
East Shewa-Guraghe	Linear	MTCI	Biweekly	0.48	1601
East Shewa-Guraghe	Linear	NDVI	Biweekly	0.33	1814
East Shewa-Guraghe	Random Forest	GCVI	Biweekly	0.44	1654
East Shewa-Guraghe	Random Forest	MTCI	Biweekly	0.52	1540
East Shewa-Guraghe	Random Forest	NDVI	Biweekly	0.37	1753
East Shewa-Guraghe	Linear	GCVI	Monthly	0.48	1606
East Shewa-Guraghe	Linear	MTCI	Monthly	0.46	1624
East Shewa-Guraghe	Linear	NDVI	Monthly	0.35	1786
East Shewa-Guraghe	Random Forest	GCVI	Monthly	0.43	1675
East Shewa-Guraghe	Random Forest	MTCI	Monthly	0.56	1475
East Shewa-Guraghe	Random Forest	NDVI	Monthly	0.29	1870
Jimma	Linear	GCVI	Monthly	0.18	1961
Jimma	Linear	MTCI	Monthly	0.29	1828
Jimma	Linear	NDVI	Monthly	0.16	1988
Jimma	Random Forest	GCVI	Monthly	0.16	1986
Jimma	Random Forest	MTCI	Monthly	0.32	1789
Jimma	Random Forest	NDVI	Monthly	0.04	2122

Table 6. R^2 and RMSE for models that were trained in one region and then applied to another region to assess generalizability.

Training Sub-Region	Validation Sub-Region	Regressor	Vegetation Index (VI)	R^2 (Training)	R^2 (Validation)	RMSE (Training)	RMSE (Validation)
East Shewa-Guraghe	Jimma	Linear	MTCI	0.47	0.17	1627	1791
East Shewa-Guraghe	Jimma	Random Forest	MTCI	0.49	0.17	1584	1791
Jimma	East Shewa-Guraghe	Linear	MTCI	0.32	0.36	1791	1736
Jimma	East Shewa-Guraghe	Random Forest	MTCI	0.35	0.30	1751	1823

We can use these models to create wall-to-wall yield maps across our study region (Figure 3). We did this by applying the beta coefficients from the best-performing linear regression model to monthly vegetation index mosaics across the study region. To ensure that we only applied our yield prediction to crop pixels, we masked out non-cropped areas using the GFSAD1000 dataset, accessed through Google Earth Engine.

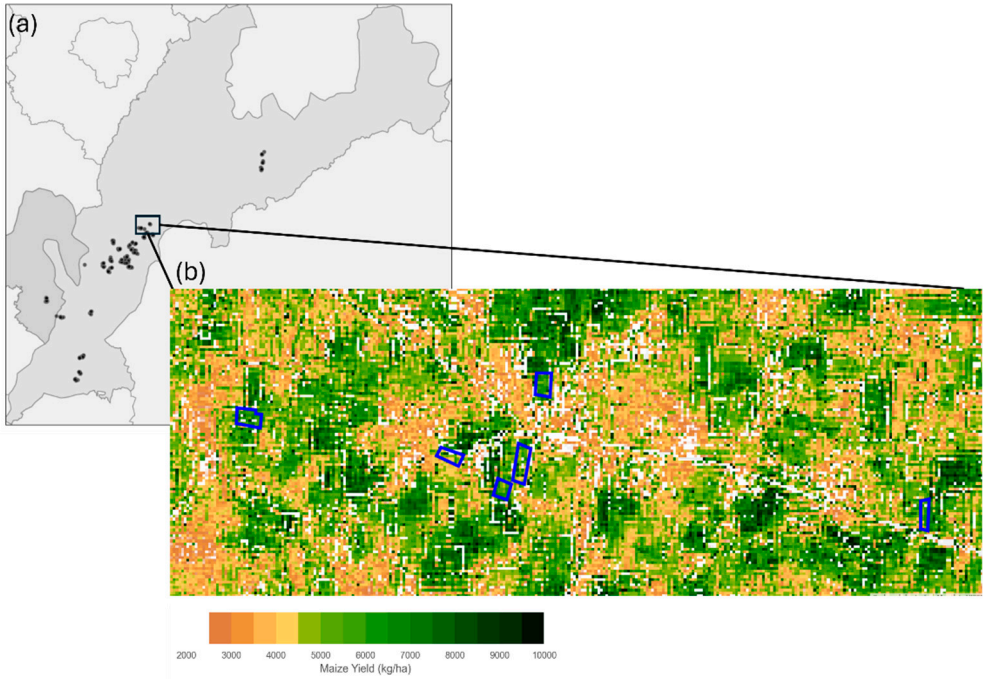

Figure 3. Pixelwise map of estimated maize yield in the (**a**) East Shewa–Guraghe sub-region. The map (**b**) uses the best-performing linear regression model with monthly MTCI predictors (R^2 = 0.47, RMSE = 1621 kg/ha. Fields from our study are highlighted in blue.

5. Discussion

This study adds to the growing body of work that assesses the use of high spatio-temporal resolution Sentinel-2 satellite data to map field-level yields in smallholder systems [17,26,57]. Using over 300 crop cut field measures of yield, we developed field-level maize yield estimates for the main growing season in two regions of Ethiopia in 2021. We examined which vegetation indices (the NDVI, GCVI, and MTCI) and which regression model (linear regression and random forest) led to the highest yield prediction accuracies. We also examined the relationship between cloud cover and model performance and assessed imputation in dealing with missing data. Furthermore, we examined whether

accuracies could be improved by adding environmental variables. Finally, we analyzed the generalizability of our yield estimation models over larger spatial scales, outside of the region in which the model was originally trained.

Overall, we were able to map yields with moderate accuracy (with R^2 values up to 0.63) (Figures 2 and 3). Our model performance was comparable to previous studies that have mapped yields in smallholder systems, which generally find R^2 values between 0.4 and 0.6 [13,25,58]. This accuracy is also comparable to the one other previous study that we know of that mapped field-level maize yields in Ethiopia [41]. Differences in model performance between the two sub-regions in our study area were notable. Models performed much better overall in the East Shewa–Guraghe sub-region than in the Jimma sub-region across all VIs and model types. This difference in accuracy could not be fully explained by the differences in the temporal availability of imagery, as we found that a linear regression model trained using monthly MTCI values performed better in the East Shewa–Guraghe sub-region (R^2 = 0.33, RMSE = 1815.2) compared to in the Jimma sub-region (R^2 = 0.24, RMSE = 1879.2, Table 3).

Considering which vegetation index led to the highest accuracy, we found that models using the MTCI had the highest accuracies. There are several reasons why the MTCI likely outperformed the other vegetation indices considered in our study. First, the MTCI and GCVI are optimized for chlorophyll detection, while the NDVI has been shown to be related to leaf area index (LAI) [50]. Previous studies have suggested that indices with increased sensitivity to chlorophyll concentrations are better able to account for the effects of nutrient stress on yields [25], which could be particularly important in our study region where fertilizer application rates were low and fields were likely nitrogen limited. Second, the NDVI may have performed poorly because it is more likely than the MTCI or GCVI to become saturated at high levels of biomass [59–63]. To support this hypothesis, we found low correlation values between the NDVI and yield compared to that of other VIs during late-season periods of peak biomass (Figures S1–S3). Third, the three vegetation indices considered in our study are differentially impacted by haze and atmospheric scattering. Specifically, the MTCI is less sensitive to haze and atmospheric effects than the GCVI and NDVI because it is computed using two nearby spectral bands that are affected similarly by atmospheric scattering [60,64]. This could be particularly important for our study region where we observed patches of haze in various images (Figure S4) despite cloud masking and image compositing. These results suggest that the MTCI performs well compared to other VIs during the rainy growing season when cloud cover and haze are extensive. Finally, we found that the MTCI still performed best despite its availability at a coarser spatial resolution compared to the other VIs considered in this study (20 m vs. 10 m). This is likely because the fields considered in our study were relatively large; the mean plot size across all fields was 3.5 ha, and only 11 fields (3% of all fields) were smaller than one hectare. Future work should examine if the MTCI still outperforms other indices in locations with very small field sizes (<0.5 ha).

We found marginal differences in model performance between linear regression and random forest regression models. Although random forest regressions led to the best-performing models across all locations (Table 3), we found them to not be as generalizable as linear regression models (Table 6). This is likely because the relationship between VIs and yield is largely linear and because random forest regressions are poor at extrapolating values outside of the range of their predictor space [65].

We found that our yield prediction accuracies were limited by cloud cover and haze. Despite our extensive cloud masking and mosaicking, there was still significantly more haze seen in imagery in the Jimma sub-region compared to the East Shewa–Guraghe sub-region (Figure S4). Excessive cloud cover can lead to missing data in addition to spectral contamination. To help overcome issues of missing data, we tested an imputation method that used a weighted moving average to fill in null values and smooth time series VI data. However, we observed decreases in model performance when imputing values, compared to simply dropping these contaminated fields from our analysis (the R^2 increased from 0.5

to 0.63, Tables 3 and 4). Guo et al. (2023) [41], also working in smallholder maize systems in Ethiopia, utilized a Savitzky–Golay filter to perform pixel-level smoothing of time series Sentinel-2 VI imagery in order to address cloud contamination. However, they observed poor results in areas with persistent cloud cover. Future work should examine whether including additional imagery that is less sensitive to cloud cover and haze, such as radar Sentinel-1 imagery, may improve yield prediction accuracies in regions plagued by high cloud cover during the rainy growing season.

The addition of environmental variables, including temperature, precipitation, and soil variables, did little to improve model performance. Previous studies have indicated that environmental variables may be able to account for spatial variability in environmental stress and their effects on yield better than vegetation indices alone [38]. The use of LST in particular has been of interest because it is sensitive to differences in soil water moisture and evapotranspiration and has thus been hypothesized to account for both water and heat-related stress [66]. We believe there are several reasons as to why the addition of environmental variables did not improve model accuracy in our study. First, we examined yield variation across a relatively small spatial extent of ~10,000 km^2 and only in one year, meaning that there was not a large amount of variation in environmental variables across our dataset. Previous studies have shown the benefit of incorporating weather variables into satellite yield prediction models at larger spatio-temporal scales where there is large variation in weather [28,36,37]. Second, we estimated yield at the field scale in a smallholder system, where the mean field size was only 3.5 hectares. Thus, it is likely that the relatively coarse spatial resolution of environmental variables in our study (250 m^2 to 5 km^2) led to issues of both mixed pixels and reduced variation across fields. Previous studies have shown the benefit of incorporating weather data into satellite yield estimation models, which mapped yield at resolutions coarser than the field scale in smallholder systems [28,38].

Considering the generalizability of our models, we found that the models trained in one region and applied to another region where no data were used for training performed poorly, particularly when applying models to the Jimma sub-region that were trained in the East Shewa–Guraghe sub-region. The poor generalizability can be attributed to differences in climatic factors, compounded by a noisy and small training dataset. The two sub-regions are located in different agroecological zones with differences in overall yields and yield variation [67]. The Jimma sub-region, which is located in a cool/humid zone, had higher observed yields on average with less variability (N = 148, mean = 6235 kg/ha, sd = 1967 kg/ha) compared to the East Shewa–Guraghe sub-region (N = 163, mean = 5461 kg/ha, sd = 2168 kg/ha), which is located in a cool/subhumid zone. Climatic differences affect model generalizability as the relationship between VIs and yield varies across space, and VIs are not always able to capture yield variability due to environmental stress [28,68]. Future work should examine how other approaches that may be more generalizable, such as those that use crop model simulations to train algorithms instead of localized ground data [27], perform when mapping yields across disparate regions. This is achieved by simulating likely crop growth (e.g., the LAI) and yield for dozens of fields in a given region, converting these simulated LAI values to satellite vegetation indices using empirically derived formulae from the literature and developing a relationship between simulated vegetation indices and yield that can be applied to satellite imagery [27].

6. Conclusions

In conclusion, we found that we were able to use Sentinel-2 satellite imagery to map field-level maize yields accurately, particularly in the eastern portion of our study region, where we achieved an R^2 of 0.63 in our best-performing model. This accuracy is similar to the best-performing models from other studies that have mapped field-level yields in smallholder systems [13,25,58]. We found that random forest models that used the MTCI were the best-performing models, though these models were less generalizable than linear regression models. We found that there were significant differences in model performance

across regions, with accuracies dropping substantially in the Jimma region which was plagued with more cloud cover and haze. We found that while imputing missing values increased the number of fields that we could consider in our analysis, it significantly reduced model prediction accuracy (a reduction in R^2 of 0.13). We also found that adding environmental variables to our models did not improve accuracy, likely because the small spatio-temporal scale of our analysis did not allow for much variation in these variables. Finally, we found that our models were not very generalizable, likely due to differences in climate, crop management, and cloud cover across our two study regions. Overall, our results highlight the strength of MTCI and Sentinel-2 for mapping field-level yields, even during the rainy season in regions with heterogeneous smallholder fields.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs16183451/s1, Figure S1: Pairwise correlation heatmap of yield and monthly NDVI composites; Figure S2: Pairwise correlation heatmap of yield and monthly GCVI composites; Figure S3: Pairwise correlation heatmap of yield and monthly MTCI composites; Figure S4: Sentinel-2A 30-day image composite.

Author Contributions: Conceptualization: M.J., A.P., J.C., T.S.S. and Z.M.; Methodology: Z.M. and M.J.; Software: Z.M.; Validation: Z.M.; Formal analysis: Z.M.; Investigation: Z.M.; Resources: M.J.; Data curation: T.S.S.; Writing—original draft preparation: Z.M.; Writing—review and editing: M.J., A.P., J.C., T.S.S. and R.W.; Visualization: Z.M.; Supervision: M.J.; Project administration: M.J.; Funding acquisition: M.J., A.P., J.C. and T.S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NASA, grant number 80NSSC21K1469.

Institutional Review Board Statement: This study was reviewed and approved as exempt by the University of Michigan Institutional Review Board (IRB): HUM00205154.

Data Availability Statement: Data to recreate yield models are available from the corresponding author, M.J. (mehajain@umich.edu).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Godfray, C.; Beddington, J.; Crute, I.; Haddad, L.; Lawrence, D.; Muir, J.; Pretty, J.; Robinson, S.; Toulmin, C. Food Security: The Challenge of Feeding 9 Billion People. *Science* **2010**, *327*, 812–823. [CrossRef] [PubMed]
2. Tilman, D.; Balzer, C.; Hill, J.; Befort, B.L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20260–20264. [CrossRef] [PubMed]
3. Lobell, D.B.; Cassman, K.G.; Field, C.B. Crop Yield Gaps: Their Importance, Magnitudes, and Causes. *Annu. Rev. Environ. Resour.* **2009**, *34*, 179–204. [CrossRef]
4. Mueller, N.; Gerber, J.S.; Johnston, M.; Ray, D.K.; Ramankutty, N.; Foley, J.A. Closing yield gaps through nutrient and water management. *Nature* **2012**, *490*, 254–257. [CrossRef]
5. Bezner Kerr, R.; Hasegawa, T.; Lasco, R.; Bhatt, I.; Deryng, D.; Farrell, A.; Gurney-Smith, H.; Ju, H.; Lluch-Cota, S.; Meza, F.; et al. Food, Fibre, and Other Ecosystem Products. In *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Pörtner, H.-O., Roberts, D.C., Tignor, M., Poloczanska, E.S., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., et al., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2022; pp. 713–906.
6. OECD/FAO. *OECD-FAO Agricultural Outlook 2022–2031*; OECD Publishing: Paris, France, 2022.
7. Di Falco, S.; Veronesi, M.; Yesuf, M. Does adaptation to climate provide food security? A micro-perspective from Ethiopia. *Am. J. Agric. Econ.* **2011**, *93*, 829–846. [CrossRef]
8. Mohamed, A. Food Security Situation in Ethiopia: A Review Study. *Int. J. Health Econ. Policy* **2017**, *2*, 86–96.
9. Abate, T.; Shiferaw, B.; Menkir, A.; Wegary, D.; Kebede, Y.; Tesfaye, K.; Kassie, M.; Bogale, G.; Tadesse, B.; Keno, T. Factors that transformed maize productivity in Ethiopia. *Food Sec.* **2015**, *7*, 965–981. [CrossRef]
10. Carletto, C.; Jolliffe, D.; Banerjee, R. From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *J. Dev. Stu.* **2013**, *51*, 133–148. [CrossRef]
11. Paliwal, A.; Jain, M. The Accuracy of Self-Reported Crop Yield Estimates and Their Ability to Train Remote Sensing Algorithms. *Front. Sustain. Food Syst.* **2020**, *4*, 25. [CrossRef]

12. Paliwal, A.; Balwinder-Singh; Poonia, S.; Jain, M. Using microsatellite data to estimate the persistence of field-level yield gaps and their drivers in smallholder systems. *Sci. Rep.* **2023**, *13*, 11170. [CrossRef]
13. Jain, M.; Srivastava, A.K.; Singh, B.; Joon, R.J.; McDonald, A.; Royal, K.; Lisaius, M.C.; Lobell, D.B. Mapping Smallholder Wheat Yields and Sowing Dates Using Micro-Satellite Data. *Remote Sens.* **2016**, *8*, 860. [CrossRef]
14. Azzari, G.; Jain, M.; Lobell, D. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sens. Environ.* **2017**, *202*, 129–141. [CrossRef]
15. Sweeney, S.; Ruseva, T.; Estes, L.; Evans, T. Mapping Cropland in Smallholder-Dominated Savannas: Integrating Remote Sensing Techniques and Probabilistic Modeling. *Remote Sens.* **2015**, *7*, 15295–15317. [CrossRef]
16. Jin, Z.; Azzari, G.; You, C.; Di Tommaso, S.; Aston, S.; Burke, M.; Lobell, D.B. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sens. Environ.* **2019**, *228*, 115–128. [CrossRef]
17. Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S. High resolution wheat yield mapping using Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111410.
18. Gao, F.; Zhang, X. Mapping crop phenology in near real-time using satellite remote sensing: Challenges and opportunities. *J. Remote Sens.* **2021**, *2021*, 8379391. [CrossRef]
19. Zhang, L.; Zhang, Z.; Luo, Y.; Cao, J.; Xie, R.; Li, S. Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using deep learning. *Agric. For. Meteorol.* **2021**, *311*, 108666. [CrossRef]
20. Aranguren, M.; Castellon, A.; Aizpurua, A. Wheat yield estimation with NDVI values using a proximal sensing tool. *Remote Sens.* **2020**, *12*, 2749. [CrossRef]
21. Johnson, D.M.; Rosales, A.; Mueller, R.; Reynolds, C.; Frantz, R.; Anyamba, A.; Pak, E.; Tucker, C. USA crop yield estimation with MODIS NDVI: Are remotely sensed models better than simple trend analysis? *Remote Sens.* **2021**, *13*, 21. [CrossRef]
22. Clevers, J.G.P.W.; Gitelson, A.A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 344–451. [CrossRef]
23. Houborg, R.; McCabe, M. A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. *Remote Sens. Environ.* **2018**, *209*, 211–226. [CrossRef]
24. Nguy-Robertson, A.L.; Peng, Y.; Gitelson, A.A.; Arkebauer, T.J.; Pimstein, A.; Herrmann, I.; Karnieli, A.; Rundquist, D.C.; Bonfil, D.J. Estimating green LAI in four crops: Potential of determining optimal spectral bands for a universal algorithm. *Agric. For. Meteorol.* **2014**, *192*, 140–148. [CrossRef]
25. Burke, M.; Lobell, D.B. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2189–2194. [CrossRef] [PubMed]
26. Jin, Z.; Azzari, G.; Burke, M.; Aston, S.; Lobell, D. Mapping smallholder yield heterogeneity at multiple scales in Eastern Africa. *Remote Sens.* **2017**, *9*, 931. [CrossRef]
27. Lobell, D.B.; Thau, D.; Seifert, C.; Engle, E.; Little, B. A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* **2015**, *164*, 324–333. [CrossRef]
28. Jain, M.; Singh, B.; Srivastava, A.A.K.; Malik, R.K.; McDonald, A.J.; Lobell, D.B. Using satellite data to identify the causes of and potential solutions for yield gaps in India's Wheat Belt. *Environ. Res. Lett.* **2017**, *12*, 094011. [CrossRef]
29. Ansarifar, J.; Wang, L.; Archontoulis, S.V. An interaction regression model for crop yield prediction. *Sci. Rep.* **2021**, *11*, 17754. [CrossRef]
30. Farmonov, N.; Amankulova, K.; Szatmári, J.; Urinov, J.; Narmanov, Z.; Nosirov, J.; Mucsi, L. Combining PlanetScope and Sentinel-2 images with environmental data for improved wheat yield estimation. *Int. J. Digit. Earth* **2023**, *16*, 847–867. [CrossRef]
31. Sibley, A.M.; Grassini, P.; Thomas, N.E.; Cassman, K.G.; Lobell, D.B. Testing Remote Sensing Approaches for Assessing Yield Variability among Maize Fields. *Agron. J.* **2014**, *106*, 24–32. [CrossRef]
32. Desloires, J.; Ienco, D.; Botrel, A. Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods. *Comput. Electron. Agric.* **2023**, *209*, 107807. [CrossRef]
33. Ray, D.; Gerber, J.; MacDonald, G.; West, P.C. Climate variation explains a third of global crop yield variability. *Nat. Commun.* **2015**, *6*, 5989. [CrossRef] [PubMed]
34. Leroux, L.; Baron, C.; Zoungrana, B.; Traoré, S.B.; Lo Seen, D.; Bégué, A. Crop Monitoring Using Vegetation And Thermal Indices For Yield Estimates: Case Study Of A Rainfed Cereal In Semi-Arid West Africa. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 347–362. [CrossRef]
35. Mladenova, I.; Bolten, J.; Crow, W.; Anderson, M.; Hain, C.; Johnson, D.; Mueller, R. Intercomparison of Soil Moisture, Evaporative Stress, and Vegetation Indices for Estimating Corn and Soybean Yields Over the U.S. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1328–1343. [CrossRef]
36. Johnson, D.M. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sens. Environ.* **2014**, *141*, 116–128. [CrossRef]
37. Johnson, D.M.; Hsieh, W.W.; Cannon,, A.J.; Davidson, A.; Bédard, F. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agric. For. Meteorol.* **2016**, *218–219*, 74–84. [CrossRef]
38. Leroux, L.; Castets, M.; Baron, C.; Escorihuela, M.J.; Begue, A.; Lo Seen, D. Maize yield estimation in West Africa from crop process-induced combinations of multi-domain remote sensing indices. *Eur. J. Agron.* **2019**, *108*, 11–26. [CrossRef]

39. Kang, Y.; Ozdogan, M.; Zhu, X.; Ye, Z.; Hain, C.; Anderson, M. Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. *Environ. Res. Lett.* **2019**, *15*, 064005. [CrossRef]
40. Debalke, D.B.; Abebe, J.T. Maize yield forecast using GIS and remote sensing in Kaffa Zone, South West Ethiopia. *Environ. Syst. Res.* **2022**, *11*, 1. [CrossRef]
41. Guo, Z.; Chamberlin, J.; You, L. Smallholder maize yield estimation using satellite data and machine learning in Ethiopia. *Crop Environ.* **2023**, *2*, 165–174. [CrossRef]
42. Hadado, T.T.; Rau, D.; Bitocchi, E.; Papa, R. Genetic diversity of barley (*Hordeum vulgare* L.) landraces from the central highlands of Ethiopia: Comparison between the Belg and Meher growing seasons using morphological traits. *Genet. Resour. Crop Evol.* **2009**, *56*, 1131–1148. [CrossRef]
43. Wakjira, M.T.; Peleg, N.; Anghileri, D.; Molnar, D.; Alamirew, T.; Six, J.; Molnar, P. Rainfall seasonality and timing: Implications for cereal crop production in Ethiopia. *Agric. For. Meteorol.* **2021**, *310*, 108633. [CrossRef]
44. Tiedeman, K.; Chamberlin, J.; Kosmowski, F.; Ayalew, H.; Sida, T.; Hijmans, R.J. Field Data Collection Methods Strongly Affect Satellite-Based Crop Yield Estimation. *Remote Sens.* **2022**, *14*, 1995. [CrossRef]
45. Gillies, S. Shapely: Manipulation and Analysis of Geometric Objects. 2007. Available online: https://github.com/Toblerity/Shapely (accessed on 3 January 2024).
46. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
47. Zupanc, A. Improving Cloud Detection with Machine Learning. Available online: https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13 (accessed on 2 January 2023).
48. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Okains with ERTS. In Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, Washington, DC, USA, 10–14 December 1973; Freden, S.C., Mercanti, E.P., Eds.; NASA: Washington, DC, USA, 1974.
49. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [CrossRef]
50. Dash, J.; Curran, P.J. The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* **2004**, *2523*, 5403–5413. [CrossRef]
51. Wan, Z. New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products. *Remote Sens. Environ.* **2008**, *112*, 59–74. [CrossRef]
52. Funk, C.; Peterson, P.; Landsfeld, M.; Pereros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data* **2015**, *2*, 150066. [CrossRef]
53. Hengl, T.; Mendes de Jesus, J.; Heuvelink, G.B.M.; Ruiperez, M.G.; Kilibarda, M.; Blagotić, A. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, 0169748. [CrossRef]
54. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
55. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
56. Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
57. Zhao, Y.; Potgieter, A.B.; Zhang, M.; Wu, B.; Hammer, G.L. Predicting wheat yield at the field scale by combining high-resolution Sentinel-2 imagery and crop modeling. *Remote Sens.* **2020**, *12*, 1024. [CrossRef]
58. Jain, M.; Balwinder-Singh; Rao, P.; Srivastava, A.K.; Poonia, S.; Blesh, J.; Azzari, G.; McDonald, A.J.; Lobell, D.B. The impact of agricultural interventions can be doubled by using satellite data. *Nat. Sustain.* **2019**, *2*, 931–934. [CrossRef]
59. Sellers, P.J. Canopy reflectance, photosynthesis and transpiration. *Int. J. Remote Sens.* **1985**, *6*, 1335–1372. [CrossRef]
60. Dash, J.; Lankester, T.; Hubbard, S.; Curran, P.J. Signal-to-noise ratio for MTCI and NDVI time series data. In Proceedings of the 2nd MERIS/(A)ATSR User Workshop, Frascati, Italy, 22–26 September 2007.
61. Gu, Y.; Wylie, B.K.; Howard, D.M.; Phuyal, K.P.; Ji, L. NDVI saturation adjustment: A new approach for improving cropland performance estimates in the Greater Platte River Basin, USA. *Ecol. Indic.* **2013**, *30*, 1–6. [CrossRef]
62. Ulfa, F.; Orton, T.G.; Dang, Y.P.; Menzies, N.W. Developing and Testing Remote-Sensing Indices to Represent within-Field Variation of Wheat Yields: Assessment of the Variation Explained by Simple Models. *Agronomy* **2022**, *12*, 384. [CrossRef]
63. Tucker, C.J. Asymptotic nature of grass canopy spectral reflectance. *Appl. Opt.* **1977**, *16*, 1151–1156. [CrossRef]
64. Lobell, D.B.; Azzari, G.; Burke, M.; Gourlay, S.; Jin, Z.; Kilic, T.; Murray, S. Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis. *Am. J. Agric. Econ.* **2020**, *102*, 202–219. [CrossRef]
65. Hengl, T.; Nussbaum, M.; Wright, N.M.; Heuvelink, G.B.M.; Graler, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* **2018**, *6*, 5518. [CrossRef]
66. Pede, T.; Mountrakis, G.; Shaw, S.B. Improving corn yield prediction across the US Corn Belt by replacing air temperature with daily MODIS land surface temperature. *Agric. For. Meteorol.* **2019**, *276*, 107615. [CrossRef]

67. Amede, T.; Auricht, C.; Boffa, J.M.; Dixon, J.; Mallawaarachchi, T.; Rukuni, M.; Deneke, T. *The Evolving Farming and Pastoral Landscapes in Ethiopia: A Farming System Framework for Investment Planning and Priority Setting*; ACIAR: Canberra, Australia, 2015; ISBN 1056875700.
68. Sakamoto, T. Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a Random Forest regression algorithm. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 208–228. [CrossRef]