

# Crop Yield Estimation of Teff (*Eragrostis tef* Zuccagni) Using Geospatial Technology and Machine Learning Algorithm in the Central Highlands of Ethiopia

Hailu Shiferaw<sup>1</sup>, Getachew Tesfaye<sup>2</sup>, Habtamu Sewnet<sup>3</sup> & Leulseged Tamene<sup>4</sup>

<sup>1</sup> Water and Land Resource Center, Addis Ababa University, P.O.Box 7985, Addis Ababa, Ethiopia

<sup>2</sup> Remote Sensing Department, Earth Observation Division, Ethiopian Space Science Technology Institute (ESSTI), Addis Ababa, Ethiopia

<sup>3</sup> Geospatial Institute, Addis Ababa, Ethiopia

<sup>4</sup> International Center for Tropical Agriculture (CIAT), Addis Ababa, Ethiopia

Correspondence: Hailu Shiferaw, Water and Land Resource Center, Addis Ababa University, P.O.Box 7985, Addis Ababa, Ethiopia. E-mail: hailushi31@yahoo.com

Received: July 13, 2021 Accepted: January 12, 2022 Online Published: January 18, 2022

doi:10.5539/sar.v11n1p34

URL: <https://doi.org/10.5539/sar.v11n1p34>

## Abstract

The genus *Eragrostis tef* Zuccagni is commonly known as “Teff”, is an indigenous cereal crop and is the major staple food crop in Ethiopia. It is mostly used to prepare a spongy flatbread called “*Injera*” and is consumed by more than 70% of the Ethiopian people. This study is conducted at nine Teff-dominated zones of the country to examine whether geospatial technology can serve to estimate the productivity of crop yield. For this, ground truth sample plots were used for nine zones, and geospatial technology and machine learning were applied for upscaling to the whole study area’s scale. Very good correlation results were obtained from spatial predictions of Teff yield for 2015 and 2020 with ROC-AUC of 89 and 91% and  $R^2$  of 0.67 and 0.73, respectively. The average predicted yields of Teff were about 1.37 t/ha and 1.99 t/ha for 2015 and 2020, respectively, indicating that such technology can offer a very good result to estimate yields for unreachable areas in the case of either during unfavorable political or other natural conditions. By doing so, we can plan to apply such technologies that can serve to save time, effort, and resources.

**Keywords:** crop yield, geospatial technology, *Injera*, machine learning, Random Forest algorithm, Teff

## 1. Introduction

The genus *Eragrostis tef* Zucc. is commonly known as “Teff” is an indigenous cereal crop of Ethiopia. Teff is one of the major crops and has the largest value in terms of both production and consumption (Nandeshwar et al., 2020; Lee, 2018; Minten et al., 2016). It is mostly used to prepare a spongy flatbread called “*Injera*” (Wato, 2019) and is consumed by more than 70% of the Ethiopian people as a staple food (Tamirat and Tilahun, 2020; Firdisa, 2016). Ethiopia is the only country that grows Teff as a major staple food crop. Teff accounts for the primary crop in area coverage and stands second in total annual production next to maize but ranks the lowest yield compared with other cereals grown in Ethiopia (Lakew and Berhanu, 2019; Tesfahun, 2018). Teff covers quite half the farmland area under cereals in Ethiopia (Habtegebrial et al., 2007), which is about 2.97 million hectares (CSA, 2014). Indeed, its production area is increasing from time to time following increased local and foreign market demands (Hailu and Seyfu, 2000). Teff is hand-broadcasted on the prepared farm field and its seeds are left uncovered (Sate and Tafese, 2016).

Besides to the seeds, farmers highly value Teff’s straw as a source of animal feed, particularly during the dry season (Lakew and Berhanu, 2019; Tesfahun, 2018; Cheng et al., 2017; Redden, 2012). The straw is also used to reinforce mud and plaster the walls of *Tukuls* for local grain storage and traditional huts (Lakew and Berhanu, 2019; Amare and Adane, 2015). Currently, it has got great attention globally due to its gluten-free crop, tolerance to biotic and abiotic stress, animal feed, and erosion control quality (Tamirat and Tilahun, 2020; Sate and Tafese, 2016; Amare and Adane, 2015). Teff is one of the strategic food crops for Ethiopia and constitutes two-thirds of the daily protein intake and 11% of the per capita caloric intake of typical households of the country (Crymes,

2015). Thus, Teff is a crucial and economically superior commodity and incorporates a significant contribution to the livelihood and food security of many Ethiopians. Looking at environmental requirements, Teff can be grown over a wide range of altitudes (i.e., from near sea level to 3,000 m.a.s.l), but the crop has shown its best performance at an altitude ranging from 1,100 and 2,950 m.a.s.l (Tadesse et al., 2016). Scholars have reported that Teff is a relatively resistant crop to several biotic and abiotic stress (FAO, 2015; Seyfu, 1997). Thus, Teff could be adapted and harvested during drought years, while other crops are failed (Zhu, 2018) when food scarcity prevails and used for food in Ethiopia.

Proper estimation of crop yield at different scales is needed to enhance the realm of food production and support the food security programs at the regional and national levels. However, the crop production estimates are often based on the conventional labor-intensive surveys, which are expensive, at risk of large errors, and aren't easily scalable and readily available on time, not comprehensive. Consequently, many more remains unknown about the quantity of crop yield (particularly Teff) produced over different locations.

Very good progresses and achievements are made in crop yield estimation using remotely sensed data (Battude et al., 2016; Johnson, 2014). Basically, the goal of remote sensing, in this case, is to spectrally measure crop biophysical variables associated to crop conditions and yield, which might subsequently be converted to actual yield estimation using different forms of deterministic or regression methods (Tesfaye et al., 2021). The utilization of satellite imagery in monitoring and estimating Teff crop yield is extremely limited both at the national and local levels. But the recent advances in remote sensing science and freely availability of high-resolution earth observation data as well as geospatial analysis algorithms have brought great opportunities to estimate the crop yield of Teff (Sewnet et al., 2021; Tesfaye et al., 2021).

Many studies reveal that remote sensing-based crop yield estimation provides reliable and accurate results than the conventional one and is now becoming a growing research area (Battude et al., 2016; Johnson, 2014). The bulk of the previous research works was focused on globally recognized cereal crops (e.g., maize, wheat, rice) cultivated over large farmlands in developed countries. However, Teff is an indigenous cereal crop of Ethiopia, which grows within the smallholder farming systems, and doesn't have international research reports.

In Ethiopia, where food production remains with subsistence farming levels, the yield estimation methods should be robust, to sufficiently signify the supply of foods and source of incomes, and should try to benefit from the advancement of high resolution data, freely available remote sensing products, and machine learning models. The main objective of this research is to estimate the yield of Teff (*Eragrostis tef*) in two time periods (2015 and 2020) using remote sensing and geospatial technologies and a robust machine learning algorithm. The specific objectives were to: (1) regress agronomic, climate, biomass, and other related factors and identify governing factors of Teff growth and production, and (2) determine the *Meher* (main) season yield of Teff using geospatial and machine learning technologies and compared with conventional estimation approach over Teff dominated areas in the nine selected zones.

## 2. Materials and Methods

### 2.1 Study Area

The study was conducted in nine Teff cluster zones named East Shewa, West Shewa, Southwest Shewa from Oromia Regional State, and East Gojjam, West Gojjam, North Shewa, South Gonder, South Wollo, and Awi zones from Amhara Regional State (Fig. 1). These Teff clustered zones were purposively selected to evaluate Teff crop yield estimation using state-of-the-art of technologies.

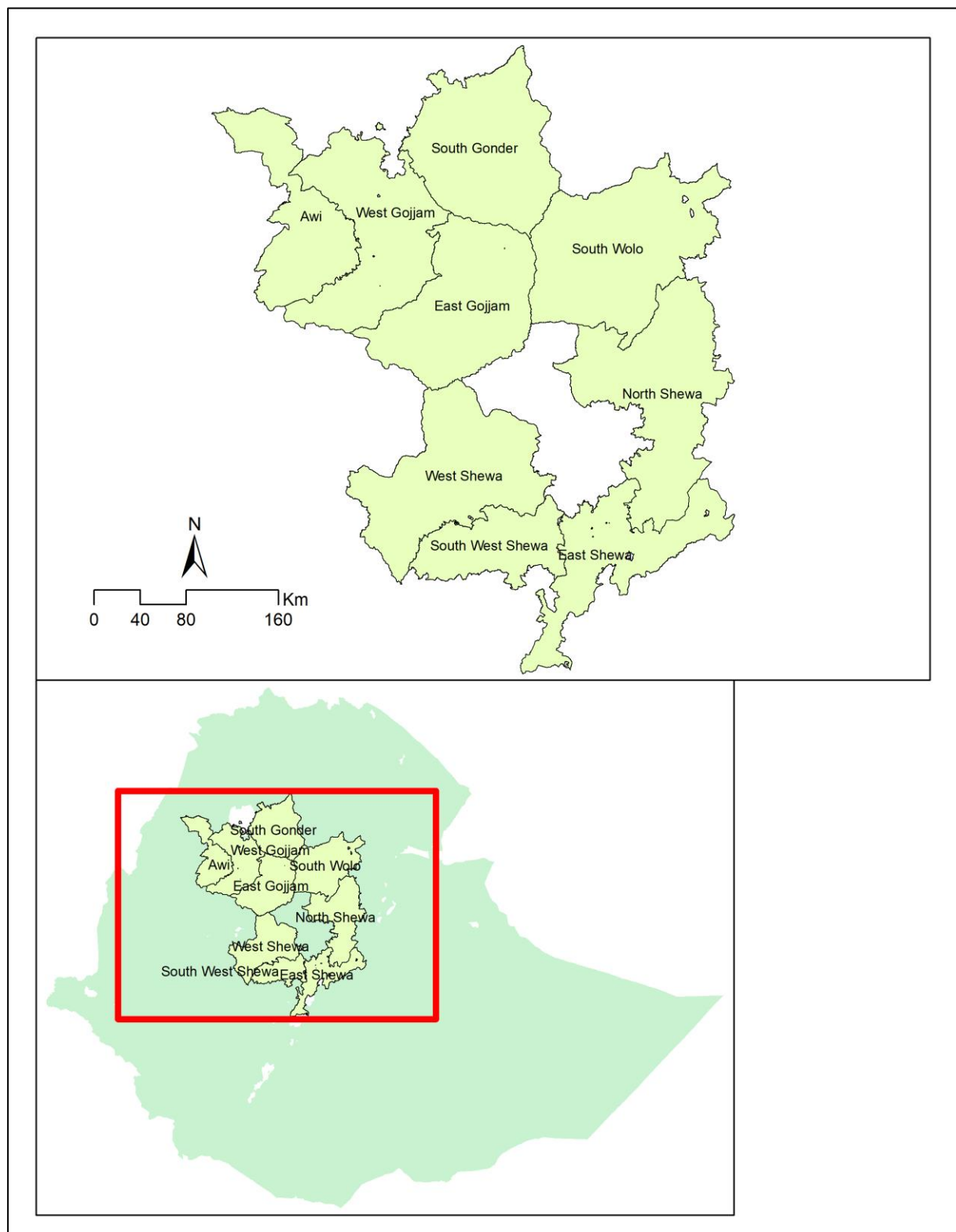


Figure 1. The study area location (the inset map is the nine zones understudy)

## 2.2 Materials

Two major groups of datasets were used for this analysis: dependent variable and independent variables. The dependent variable is the Teff yield collected during *Meher* (main) season, which is ground truth data used for calibration and validation. *Meher* season's Teff yield statistics, collected through Agricultural Sample Survey

(AgSS) based on private peasant holdings in rural areas of the country, which were obtained from Central Statistical Agency (CSA) for 2015 and 2020 time periods. Similar periods of explanatory variables were used for this analysis. The dataset includes ground truth sample plots for 2015 and 2020 collected from nine selected Teff producing zones.

On the other hand, independent variables commonly called explanatory variables were acquired from different sources. About 20 variables were used and grouped into four major types: (1) Vegetation indices derived from MODIS images: normalized difference vegetation index (NDVI), leaf area index (LAI), fraction of photosynthesis active radiation (fPAR), net primary productivity (NPP), and soil adjusted vegetation index (SAVI) were obtained from earth explorer (<https://earthexplorer.usgs.gov/>); (2) Climatic data: mean monthly rainfall and temperature, length of growing periods (LGP) were obtained from Climate Hazards Group Infra-Red Precipitation with Station data (CHIRPS, <https://www.chc.ucsb.edu/data/chirps>); (3) Soil fertility: soil organic carbon (SOC) and organic matter (OM), and (4) Topographic factors: compound topographic index (CTI), digital elevation model (DEM) and slope of the study site.

### 2.3 Data Set Preparation and Approach

Datasets were acquired from different sources. Resampling into 30 m pixel sizes, re-projection from sinusoidal system to a local coordinate system (WGS\_1984\_UTM\_ZONE\_37), clipping into the study zones, indices calculations, and other necessary preparations were performed to arrange input datasets corresponding to the study area. Furthermore, machine learning (e.g., random forest - RF) based open source and a stand-alone system were employed for analysis. The analysis was carried out in an open-source R programming environment (R Core Team, 2019) and mapping was done using an open-source of QGIS3.8.

### 2.4 Non-parametric Machine Learning Based for Teff Yield Prediction

This part was planned to predict the yield of Teff using ground truth and RS data in a non-parametric machine learning algorithm approach. Random forest, a non-parametric machine learning algorithms were used to predict Teff yield at the selected Teff clustered areas. Teff biomass estimated from earth observation, climatic data, and agronomic factors were supposed to be used as input predictor variables whereas sampled Teff yield was used for response variable to train and validate the model developed.

Random Forest package by Laiw and Wiener (2002) in R (CRAN) was used to undertake Teff yield prediction using multivariate regression (R Core Team, 2019). RF is an algorithm for classification and regression problems and predicts well when there is missing data, avoids over-fitting problems, produces more stable results, and is less sensitive to multi-collinearity than other machine learning algorithms (e.g., Classification and Regression Tree (CART) (Lai and Tsai, 2019; Fang et al., 2020; Chen et al., 2020). RF, therefore, works by growing a large collection of de-correlated decision trees as a base learner using a fraction of observation and features (variables) randomly selected with replacement (bootstrapping). Each tree (regression) was trained using 80% of randomly selected model calibration with the remaining 20% of samples were used for validation using a method called out-of-bag (OOB) samples, serving to estimate the regression (Shiferaw et al., 2019). Finally, the majority voting or mode rule (Ghimire et al., 2010) is used to assign a pixel to a class based on the maximum number of votes that the pixel receives from the group of regression trees (Breiman, 2001). A generic non-parametric regression in the machine learning equation is presented to show how the function is implemented (equation 1).

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad (1)$$

Where,  $\theta$  is what represents the current weights of predictors, and  $J(\theta)$  means the 'response for current weights',  $m$  is the number of observations,  $y$  is the real value from ground truth data,  $h_{\theta}$  is the expected value for that predictor  $x_i$  (Geitgey, 2014; URL1, 2017).

Random forest (RF) algorithm does not need for prior data transformation or elimination of outliers but can be fitted complex nonlinear relationships (Elith et al., 2008). That is, these algorithms automatically handle interaction effects among predictors (Elith et al., 2008; Breiman, 2001). RF handles both classifications and regressions, which is an enhancement of traditional decision trees by consisting of many trees as a predictor (Breiman, 2001). Each tree votes for its preferred class and the most voted class gives the final prediction (Lorena et al., 2011).

RF was found a highly performing state-of-the-art machine learning algorithm based on an ensemble of decision trees, which compliments the findings of Rembold et al. (2015). RF has several benefits compared to traditional

classifiers (maximum likelihood, for example) according to different researchers (Hastie et al., 2009; Gislason et al., 2006; Breiman, 2001). Some of its advantages are: (1) it can handle thousands of input variables and identify most significant variables so that it is considered one of the dimensionality reduction methods and also identifies importance of variables from the targeted variables, which can be a very handy feature, (2) it has an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing (“out-of-bag or bagging” system), (3) normal distribution or unimodality assumption doesn’t necessary and being relatively insensitive to the number and multi-collinearity of input data (Elith et al., 2008), (4) handles high-dimensional data (many predictors), (5) it handles categorical data together with metric (numeric) data, (6) it is very efficient and reliable, mostly no need to cross-validation (Breiman, 2001), though this study had to produce validation results to produce reliable results for our purpose, and (7) no “overfitting”, meaning no modelling noisy data since it uses tree branch pruning methods as a form of cross-validation (Breiman, 2001). However, the challenge of using RF requires expertise to handle programming of machine learning.

About one-fifth (20%) of the overall sample is left for validation (the out-of-bag predictions – OOB) while four-fifth (80%) was used for model calibration (Shiferaw et al., 2019). Each split of the tree is determined using a randomized subset of the predictors at each node and the final result is the average of the results of all the trees (Breiman, 2001). The number of features was set to the square root of the number of input variables, as has been done in another study (Belgiu and Drăgu, 2016). The OOB error was carefully checked for its stability with the chosen settings before applying the model.

### 2.5 Important/influencing Variables

Teff’s growth and production are determined by different factors such as biotic and abiotic ones. Some predictor or influencing variables were selected with expert knowledge first and then tested their responses in non-parametric analysis. Those predictors were assumed to be very important factors to explain the dependent variable (yield). Important variables are predictors, which have relatively higher influencing weight than their batches. Natekin and Knoll (2013) defined the influence of the variable  $j$  in a single tree  $T$ . Considering that the tree has  $L$  splits, all the non-terminal nodes from the root to the  $L-1$  level of the tree (Natekin and Knoll, 2013) (equation 2):

$$\text{Influence}_j(T) = \sum_{i=1}^{L-1} I_i^2 1(S_i = j) \quad (2)$$

This measure is based on the number of times a variable is selected for splitting, i.e., the current splitting variable  $S_i$  is the same as the queried variable  $j$ . The measure also captures weights of the influence with the empirical squared improvement  $I_i^2$ , assigned to the model as a result of this split (Natekin and Knoll, 2013). To obtain the overall influence of the variable  $j$  in the model, this influence should be averaged overall boosted trees (as indicated in equation 3). Accordingly, the resulting influences can then be used for both forward and backward feature selection procedures.

$$\text{Influence}_j = \frac{1}{M} \sum_{i=1}^M \text{Influence}_j(T_i) \quad (3)$$

Finally, using highly influencing predictor variables and ground truth datasets, the Teff yield and area are then generated as a continuous layer. Furthermore, to quantify and calculate the Teff growing area, a threshold level was applied and masked out the Teff growing areas within the selected nine zones.

### 2.6 Model Performance

Model performance was assessed by calculating its accuracy based on the root-mean-square error (RMSE) and coefficient of determination ( $R^2$ ), and the Receiver Operating Characteristics (ROC) of the area under the curve (AUC). The model performance was assessed by calculating its accuracy, coefficient of determination, and the Receiver Operating Characteristics (ROC) of the area under the curve (AUC). Moreover, we calculated the sensitivity (true positive rate - TPR) and specificity (true negative rate - TNR) of the model (Metz, 1978; Liaw and Wiener, 2002). This approach identified the minimum area where Teff is growing whilst ensuring that no localities at which the Teff has been omitted within the selected zones, i.e. omission rate of the minimum (Pearson, 2010) with a certain precision obtained from the model.

### 3. Results

#### 3.1 Predictors and Predicted Teff Yield

Random Forest results indicate that the best mtry (mtry is the number of variables randomly sampled as candidates at each split) was 2 of the other two mtry (11 and 20). Among the predictor variables, SAVI, precipitation in March, and CTI are the highest predictors (Fig. 2). The validation results indicate that the random forest algorithm achieved very good results with ROC-AUC of 89% and 91% for 2015 and 2020, respectively, and with  $R^2$  of 0.67 and 0.73, respectively, with the smallest RMSE values of 2.422204 and  $R^2$  of 0.44.

Considering important variables for Teff productivity, SAVI, Precipitation in July and February, SOC and CTI are the highest contributors (Fig. 2). SAVI is a soil-adjusted vegetation index, indicating the healthier environment of the area with good vegetation growth that supports potential yield. Precipitation obtained in July and February in some areas is very critical for Teff production. Rainfall in July in many parts of the country, particularly for the main season (*Meher*), and February and March precipitation is also very important for the commonly known “*Belg*” season. The “*Belg*” rainfall is mainly important for the western part of the study area (e.g. West Shewa and Southwest Shewa zones) and these areas are partly supported by “*Belg*” season productions. SOC is another very important variable that is mainly obtained from crop residue and additional organic matter applications. However, organic matter applications are not widely used for vast areas of Teff production but crop residue from previous harvest or fallow practice that support other plant species grown and produce organic matters are main sources of SOC in crop production. Hence, if the smallholder farmers either leave crop residue uncollected for fire fuel or animal feed or fallow their farmland, they will be obtained more yield from their farmland. CTI is the compound topographic index, which indicates the feature of the landscape that has highly correlation with soil moisture-holding capacity.

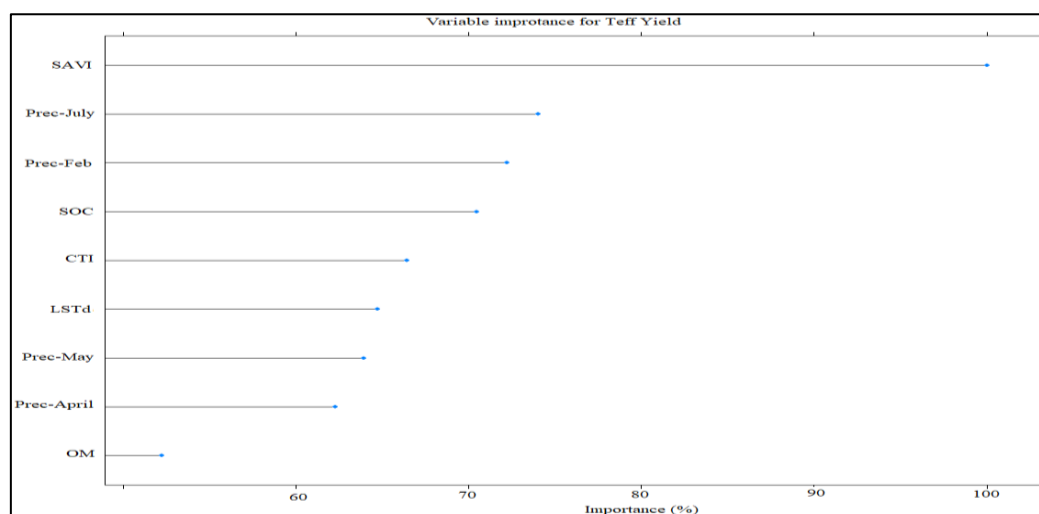


Figure 2. Important variables to predict Teff yield (x-axis is variable importance in percent)

#### 3.2 Upscaling to the Study Areas from Samples

The spatially up-scaled Teff yield probabilities in 2015 and 2020 indicate that there is little change or increase of Teff yield between 2015 and 2020 that the maximum was 1.52 t/ha in 2015 but it was increased to 2.42 t/ha in 2020. Not only the maximum probability of productivity but the minimum was increased from 1.17 t/ha in 2015 to 1.36 t/ha in 2020 (Fig. 3). These changes are encouraging and indicating that there are huge potentials for higher productivity as it is observed that there are positive changes, particularly in the southern and eastern parts of the study area, i.e., North Shewa and South Wollo of Amhara Region. The average productivity of Teff is increased from 1.37 t/ha in 2015 to 1.99 t/ha in 2020, indicating that there are considerable changes over time though it is still below its potential as these areas as naturally suitable and known as Teff belt zones of the country.



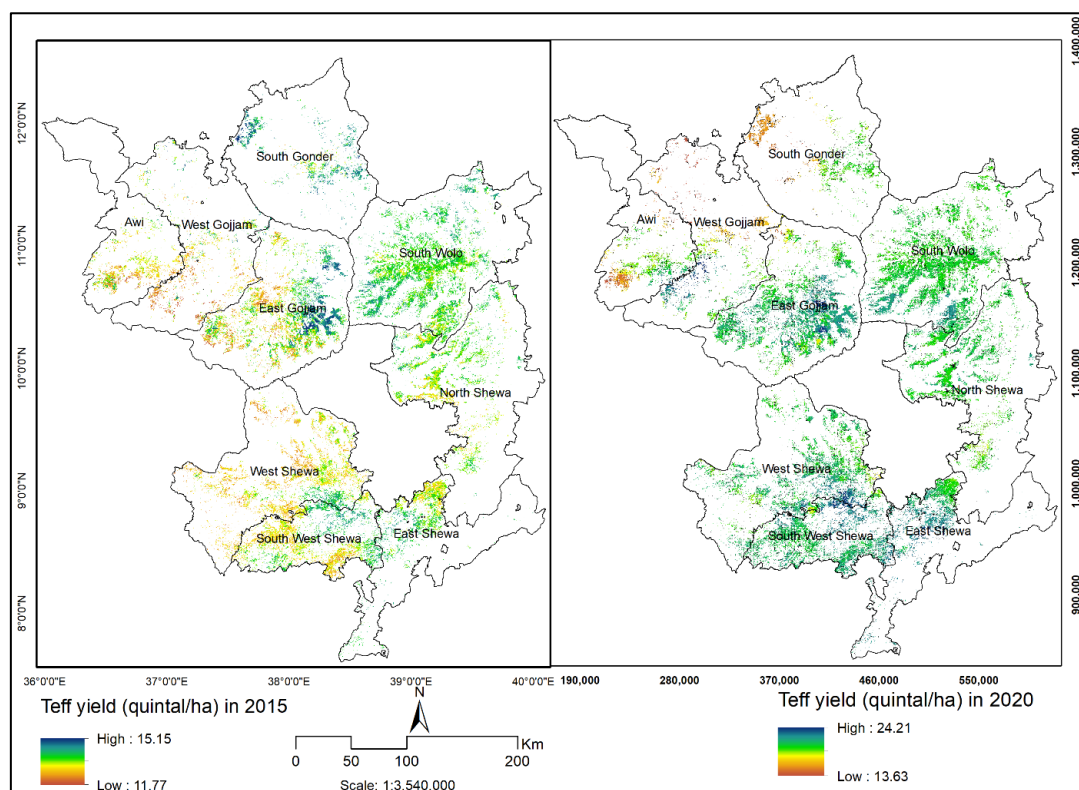


Figure 3. Teff yield in 2015 (left) and 2020 (right)

Among the nine Teff producing zones considered in this analysis, East Gojjam, West Gojjam and North Shewa of Amhara Region and East Shewa and South West Shewa of Oromia Region showed the highest Teff producing zones in both periods (Table 1). Of these, East Gojjam (Amhara Region) and South West Shewa (Oromia Region) recorded the highest Teff yield (productivity) and production in 2020.

Table 1. Teff yield prediction results summarized by zonal statistics of zone values for 2015 and 2020 using geospatial technology

Zone	Teff Yield in 2015				Teff Yield in 2020			
Name	Min	Max	Mean	Std	Min	Max	Mean	Std
Awi	12.30	14.67	13.50	0.313	15.18	15.82	15.5	0.321
East Gojjam	11.60	<b>15.15</b>	13.65	0.436	22.21	<b>23.09</b>	<b>22.65</b>	0.435
North Shewa-Amhara	12.80	14.72	13.77	0.246	20.79	21.33	<b>21.06</b>	0.274
South Gondar	12.58	15.07	13.99	0.330	18.29	18.93	18.61	0.320
South Wolo	12.88	14.74	13.89	0.235	17.12	17.56	17.34	0.221
West Gojjam	11.48	<b>15.14</b>	13.48	0.515	20.00	20.80	<b>20.4</b>	0.398
East Shewa	12.87	14.56	13.90	0.213	20.82	21.38	<b>21.1</b>	0.281
South West Shewa	12.74	14.57	13.66	0.320	20.97	<b>21.55</b>	<b>21.26</b>	0.290
West Shewa	12.47	14.55	13.38	0.279	20.57	21.29	<b>20.93</b>	0.361
Average	12.41	<b>14.83</b>	13.69	0.32	19.55	20.19	<b>19.87</b>	0.32

#### 4. Discussion

Proper estimation or forecasting of Teff crop yield at different scales is needed to enhance the domain of food production and support to address food security at the regional and national levels (Sewnet et al., 2021). However, the crop production estimates are commonly dependent on the conventional labor-intensive surveys, which are expensive, at risk of large errors, and do not easily scalable and not readily available on time, and not comprehensive (CSA, 2016). Consequently, much remains unknown about the quantity of Teff crop yield produced over different locations from where samples were not collected (Sewnet et al., 2021).

Many complex factors are contributing to Teff's productivity and its prediction over space. For example, crop

biomass plays an important role in the global carbon cycle, which is also closely related to crop conditions and yield (Zhang et al., 2013). Biomass can be used for estimating yield even at high spatial resolution with high fragmented smallholders' plots. Yield prediction in precision farming is considered of high importance for the improvement of crop management and market planning. Once the yield is site-specifically predicted, the farm inputs such as fertilizers could be applied variably according to the expected crop, yield, and soil needs (Pantazi et al., 2016).

On the other hand, the amount of carbon available as plant materials is associated to crop yield (Tao et al., 2005), implies the variation in the accumulation of biomass for Teff crops growing at different sites of the study area (Tefaye et al., 2021). The rate of accumulation of organic carbon and its storage in plants are affected by many factors including solar energy input, temperature, available soil moisture, the level of carbon dioxide, nutrient availability, the potential of the crop variety, crop management practices, among others (Hay, 1995; Sinclair and Muchow, 1999). All these factors affect the crop's light use efficiency, which is an important indicator of crop photosynthesis and depicts the efficiency with which crop to produce good yield (Ma et al., 2020).

The yield of Teff in 2020 was ranged from 1.36 to 2.41 t/ha with an average of 1.99 t/ha, indicating that it was very good productivity as compared to the previous results recorded in many parts of the country with an average yield of 1.48 t/ha (CSA, 2016). Our prediction results are also within the same ranges of productivity of Teff reported by CSA (CSA, 2020). This indicates that a very good yield estimation is possible using geospatial technologies with some reference ground truth data, with minimized cost and time that would take longer periods and higher costs for ground survey especially for those larger crop types of the country such as Teff, wheat, and maize.

However, the productivity of Teff is still very low as compared to some experiments conducted, for example by Tesfahun (2018), which is about 3.77 t/ha with the application of NPS fertilizer. Some of the limiting factors contributing to the low yield of Teff are low soil fertility, suboptimal use of mineral fertilizers, weeds, uneven rainfall distribution in lower altitudes, lack of high yielding cultivars, lodging, water-logging, and low moisture (Fenta, 2018; Tesfahun, 2018; Tamirat and Tilahun, 2020). Furthermore, random broadcasting of Teff sowing also contributes to the low productivity of Teff due to self-competition in space and minerals. Bekalu and Tenaw (2015) indicated that the most common way of planting Teff is by broadcasting the small seed at the rate of 25-50 kg/ha. This broadcasting reduces the amount of grain production, promotes competition among plants for resources, and causes severe lodging, which is the main cause for the low yield of Teff due to high plant density (Hundera et al., 2011). Using row planting or transplanting technology, weeding can be done easily and the lodging incidence can be reduced (Sebsebe and Assefa, 2013). Thus, it is important to develop an appropriate sowing method and balanced blended fertilizer application for enhancing the productivity of the crop and food security (Tefahun, 2018). However, limited research has been done to elucidate the response of Teff variety to blended fertilizer applications and row planting technologies.

Furthermore, geospatial technologies and machine learning algorithms offer very good results in predicting the production and productivity of Teff with some reasonable ground truth samples. This has many advantages: (1) estimating productivity and then food security in time, (2) plan any immediate interventions in terms of agricultural produce including Teff productivity, and (3) the technology reduces cost, saves time and resources for yield estimation using conventional methods of cut and weight systems.

## 5. Summary

It is also possible to estimate crop yield and monitor crop growth using geospatial and remote sensing technology for clustered crop areas. This approach reduces costs and efforts but requires ground truth data, technology, and skill. Crop yield estimation and growth monitoring help to plan for growers, government, and crop insurance companies and contribute to the national goal of food security. However, there are some caveats for this kind of study: (1) some of the areas lack local level ground truth data for several years, (2) the approach requires further regression analyses based on selected machine learning algorithms, (3) the future analyses are based on many variables including crop factors such as climate, agronomic factors, crop factors (e.g., harvest index), among others, besides to ground truth, and (4) our assumption was to test yield estimation potential using geospatial technology in combination with machine learning algorithms at clustered areas but this requires to test further for its applicability for fragmented cropping systems.

## Acknowledgment

The authors acknowledge the initiations and encouragement of coalition of the willing (CoW), particularly Dr. Lulseged Tamene from the International Center for Tropical Agriculture (CIAT) for supporting us to set up different team members in identifying major bottlenecks of the country that can be solved with the advancement



of geospatial technology. We acknowledge the support of CIAT to facilitate the logistics to undertake this assignment. We also acknowledge the Ethiopian Central Statistical Agency (CSA) for providing crop yield data used for the study.

## References

- Amare, A., & Adane, L. (2015). *Determination of seed rate and variety on the growth and yield of Tef in the Eastern Amhara region*. A Report, Ethiopia.
- Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., ... Demarez, V. (2016). Estimating Maize Biomass and Yield over Large Areas Using High Spatial and Temporal Resolution Sentinel-2 like Remote Sensing Data. *Remote Sensing of Environment*, 184, 668-681. <http://dx.doi.org/10.1016/j.rse.2016.07.030>
- Breiman, L. (2001). *Random forest 2001*. pp. 1-33. <https://doi.org/10.1017/CBO9781107415324.004>
- Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens*, 114, 24-31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Cheng, A., Mayes, S., Dalle, G., Demissew, S., & Massawe, F. (2017). Diversifying crops for food and nutrition security - A case of Tef. *Biological Reviews*, 92(1), 188-198. <https://doi.org/10.1111/brv.12225>
- Crymes, R. (2015). The International Footprint of Teff: Resurgence of an Ancient Ethiopian Grain. *Arts & Sciences Electronic Theses and Dissertations*, 394. Retrieved from [https://openscholarship.wustl.edu/art\\_sci\\_etds/394](https://openscholarship.wustl.edu/art_sci_etds/394)
- CSA. (2014). *Agricultural sample survey report on area and production for major crops*. Private Peasant Holdings Meher Season. The FDRE Statistical Bulletin, General information about Teff in Ethiopia. Addis Ababa, Ethiopia,
- CSA. (2020). *Agricultural sample survey report on area and production for major crops*. Private Peasant Holdings Meher Season. Addis Ababa, Ethiopia.
- FAO. (2015). *Analysis of price incentives for Teff in Ethiopia Technical notes series*. MAFAP, produced by Assefa B. Demeke M., Lanos B, Rome.
- Fenta, A. (2018). Effect of Teff variety and rates of nitrogen fertilizer application on growth and yield components under Jimma condition. *ARPJ Journal of Agricultural and Biological Science*, 13(3), 42-43.
- Firdisa, B. (2016). Determinants of smallholder farmers' participation decision in tef production: evidence from horo and Jimma geneti woreda, Ethiopia. *Developing Country Studies*, 6(10), 36-43.
- Geitgey, A. (2014). *Machine Learning is Fun!* The world's easiest introduction to Machine Learning. Retrieved from <https://medium.com/@ageitgey>
- Habtegebrail, K., Singh, B. R., & Haile, M. (2007). Impact of tillage and nitrogen fertilization on yield, nitrogen use efficiency of Teff (*Eragrostis tef*) and soil properties. *Journal of Soil Tillage Research*, 94, 56-63. <https://doi.org/10.1016/j.still.2006.07.002>
- Hailu, T., & Seyfu, K. (2000). Production and Importance of Teff in Ethiopian Agriculture. In T. Hailu, B. Getachew & M. Sorrells (Eds.). *Narrowing the rift: Teff research and development* (pp. 3-7). Proceedings of the international work shop on Teff Genetics and Improvement, Debre Zeit, Ethiopia.
- Hay, M. (1995). Harvest index: a review of its use in plant breeding and crop physiology. *Ann Appl Biol*, 126, 197-216. <https://doi.org/10.1111/j.1744-7348.1995.tb05015.x>
- Hundera, F., Behute, B., Simons, R., & Berhe. T. (2011). *Tef Diagnostic Report: Strengthening the Tef Value Chain in Ethiopia*. Addis Ababa, Ethiopia. <http://dx.doi.org/10.1111/j.1601-5223.2001.00103.x>
- Johnson, D. (2014). An Assessment of Pre-and Within-Season Remotely Sensed Variables for Forecasting Corn and Soybean Yields in the United States. *Remote Sensing of Environment*, 141, 116-128. <https://doi.org/10.1016/j.rse.2013.10.027>
- Kemanian, A. R., Stockle, C., Huggins, D., & Viegas, L. (2007). A simple method to estimate harvest index in grain crops. *Field Crops Research*, 103, 208-216. <https://doi.org/10.1016/j.fcr.2007.06.007>
- Lakew, A., & Berhanu, T. (2019). Determination of seeding rate and inter-row spacing on the yield of tef (*Eragrostis tef* Zucc. Trotter) in the dryland areas of Wag Lasta, North Eastern Amhara, Ethiopia. *Archives of Agriculture and Environmental Science*, 4(1), 69-74. <https://doi.org/10.26832/24566632.2019.040101>
- Lee, H. (2018). Teff, A rising global crop: current status of teff production and value chain. *The Open Agriculture*,

- 12(1), 185-193. <https://doi.org/10.2174/1874331501812010185>
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R news*, 2, 18-22.
- Ma, L., Wang, S., Chen, J., Chen, B., Zhang, L., Ma, L., Amir, M., Sun, L., Zhou, G., & Meng, Z. (2020). Relationship between Light Use Efficiency and Photochemical Reflectance Index Corrected Using a BRDF Model at a Subtropical Mixed Forest. *Remote Sens.*, 12, 550. <https://doi.org/10.3390/rs12030550>
- Metz, C. E. (1978). Basic principles of ROC analysis. *Semin. Nucl. Med.*, 8, 283-298. [https://doi.org/10.1016/S0001-2998\(78\)80014-2](https://doi.org/10.1016/S0001-2998(78)80014-2)
- Minten, B., Tamru, S., Engida, E., & Kuma, T. (2016). Feeding Africa's cities: The case of the supply chain of Teff to Addis Ababa. *Econ Dev Cult Change*, 64(2), 265-97. <https://doi.org/10.1086/683843>
- Nandeshwar, B. C., Hirko, O., Morketa, G., Belay, G., Usman, M., Reta, F., Fufa, M., & Kinde, L. (2020). Teff (*Eragrostis tef* [Zucc] Trotter): an emerging global demanding crop. *Agriculture Observer*, 1(1), 28-33.
- Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., & Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, 121, 57-65. <https://doi.org/10.1016/j.compag.2015.11.018>
- Pearson, R. G. (2010). Species' distribution modeling for conservation educators and practitioners. *Lessons Conserv*, 3, 54-89.
- R Core Team. (2019). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org>
- Sate, S., & Tafese, A. (2016). Effects of sowing methods and seed rates on yield components and yield of tef in Soro Woreda, Hadya Zone, Southern Ethiopia. *Journal of Natural Sciences Research*, 6(19), 109-114.
- Sebsebe, C., & Assefa, K. (2013). *The agronomy of teff*. Improved evidence towards better policies for the teff value chain conference, Addis Ababa, Ethiopia. <http://dx.doi.org/10.2147/idr.s134369>
- Sewnet, H., Tesfaye, G., Shiferaw, H., & Tamene, L. (2021). *Phenology based Time Series LAI as a Proxy for Teff Crop Yield Estimation: A Case in Major Teff (Eragrostis) Growing Zones of Ethiopia*. pp. 30.
- Seyfu, K. (1997). *Teff. Eragrostis teff (Zucc.) Trotter. Promoting the conservation and use of underutilized and neglected crops*. Institute of plant genetics and crop plant research. Rome, Italy: Gatersleben / International Plant Genetic Resources Institute.
- Shiferaw, H., Schafner, U., Bewket, W., Alamirew, T., Zeleke, G., Teketay, D., & Eckert, S. (2019). Modelling the current fractional cover of an invasive alien plant and drivers of its invasion in a dryland ecosystem. *Scientific Reports*, 9, 1576. <https://doi.org/10.1038/s41598-018-36587-7>
- Sinclair, T. R., & Muchow, C. (1999). Radiation use efficiency. *Adv. Agron.*, 65, 215-265. [https://doi.org/10.1016/S0065-2113\(08\)60914-1](https://doi.org/10.1016/S0065-2113(08)60914-1)
- Tadesse, D., Alem, T., Wossen, T. (2016). Evaluation of improved varieties of Teff in West Belesa, Northwest Ethiopia. *Review of plant studies*, 3, 1-6. <https://doi.org/10.18488/journal.69/2016.3.1/69.1.1.6>
- Tamirat, W., & Tilahun, N. (2020). The response of Teff [*Eragrostis teff* (Zucc.) trotter] to nitrogen fertilizer and development of Barley (*Hordeum vulgare* L.) at Bore district, Southern Oromia, Ethiopia. *American Journal of Life Sciences*, 2(5), 260-266. <https://doi.org/10.11648/j.ajls.20140205.12>
- Tesfahun, W. (2018). Tef yield response to NPS fertilizer and methods of sowing in East Shewa, Ethiopia. *Journal of Agricultural Sciences - Sri Lanka*, 13(2), 162-173. <https://doi.org/10.4038/jas.v13i2.8340>
- Tesfaye, G., Sewnet, H., Shiferaw, H., Desta, L., Abera, W., & Gudeta, K. (2021). *Tef [Eragrostis tef (Zucc.) Trotter] yield estimation based on satellite-derived light use efficiency model*. A case study in the top tef growing zones of Ethiopia. pp. 38.
- Tao, F., Yokozawa, M., Zhang, Z., Xu, Y., & Hayashi, Y. (2005). Remote sensing of crop production in China by production efficiency models: models comparisons, estimates and uncertainties. *Ecological Modelling*, 183, 385-396. <https://doi.org/10.1016/j.ecolmodel.2004.08.023>
- van Herwaarden, A. F., Farquhar, G. D., Angus, J. F., Richards, R. A., & Howe, G. N. (1998). 'Haying-off', the negative grain yield response of dryland wheat to nitrogen fertilizer. I. Biomass, grain yield, and water use. *Aust. J. Agric. Res.*, 49, 1067-1081. <https://doi.org/10.1071/A97039>
- Wato, T. (2021). Tef [*Eragrostis tef* (Zucc)] grain yield response to nitrogen fertilizer rates in East Badewacho

- district, Hadiya Zone, Southern Ethiopia. *Cogent Food & Agriculture*, 7, 1-11.  
<https://doi.org/10.1080/23311932.2021.1909203>
- Wato, T. (2019). Effects of nitrogen fertilizer rate and inter-row spacing on yield and yield components of teff [*Eragrostis teff* (Zucc.) Trotter] in Limo district, Southern Ethiopia. *International Journal of Plant & Soil Science*, 31(3), 1-12. <https://doi.org/10.9734/IJPSS/2019/v31i330211>
- Zhang, M., Wu, B., Zeng, H., & Xing, Q. (2013). *Crop estimation by integration of remote sensing and meteorological data*. Institute of Remote Sensing and Digital Earth, Chinese Academy of Science.
- Zhu, F. (2018). Chemical composition and food uses of Teff (*Eragrostis tef*). *Food Chem.*, 239, 402-15.  
<https://doi.org/10.1016/j.foodchem.2017.06.101>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).