Real-Time Prediction of Crop Yields From MODIS Relative Vegetation Health: A Continent-Wide Analysis of Africa

Lillian Kay Petersen

Abstract: Developing countries often have poor monitoring of weather and crop health, leading to slow responses to droughts and food shortages. I developed satellite analysis methods and software tools to predict crop yields three months before the harvest. This software measures relative vegetation health based on pixel-level vegetation indices (VIs). VIs are a measure of plant health that are calculated from the light spectrum emitted from the land. Because this method requires no crop mask or subnational yield data, it can be applied to any crop or climate, making it ideal for African countries with small fields and poor ground observations. A validation was first conducted in Illinois where there is reliable county-level crop yield data. The monthly VIs were extremely well correlated with corn, soybean, and rice yields, showing that this model has good forecasting skill for crop yields. Next, the vegetation health was measured in every country in Africa to predict crop yields for the 2018 harvest. The yield predictions were very accurate with a median error of 8.6%. This method is unique because of its simplicity and versatility: it shows that a single user can produce reasonable real-time estimates of crop yields across an entire continent. For more details, see: Petersen, L.K. Real-Time Prediction of Crop Yields From MODIS Relative Vegetation Health: A Continent-Wide Analysis of Africa. Remote Sens. 2018, 10, 1726., https://www.mdpi.com/359918

---

1. Introduction

In the United States and Europe, there is exceptional monitoring and reporting of weather and crop health. With thousands of weather stations and regional crop yield data [1,2], crop yields may be predicted based on historical records [3]. However, not all parts of the world have open, reliable data [4]. Lack of data is a particularly important problem in developing countries where crop yields are less stable and droughts can lead to famines, death, government instability, and war.

Recent years have shown an advancement in strategies to obtain better data coverage in developing countries [5]. For example, detailed surveys offer researchers insights into African agriculture. However, these methods require expensive ground-based operations and remain difficult to scale across a large area, and are often quite inaccurate [6]. Agriculture is one of the backbones of African economies and provides food, income, power, stability, and resilience to rural livelihoods [7].

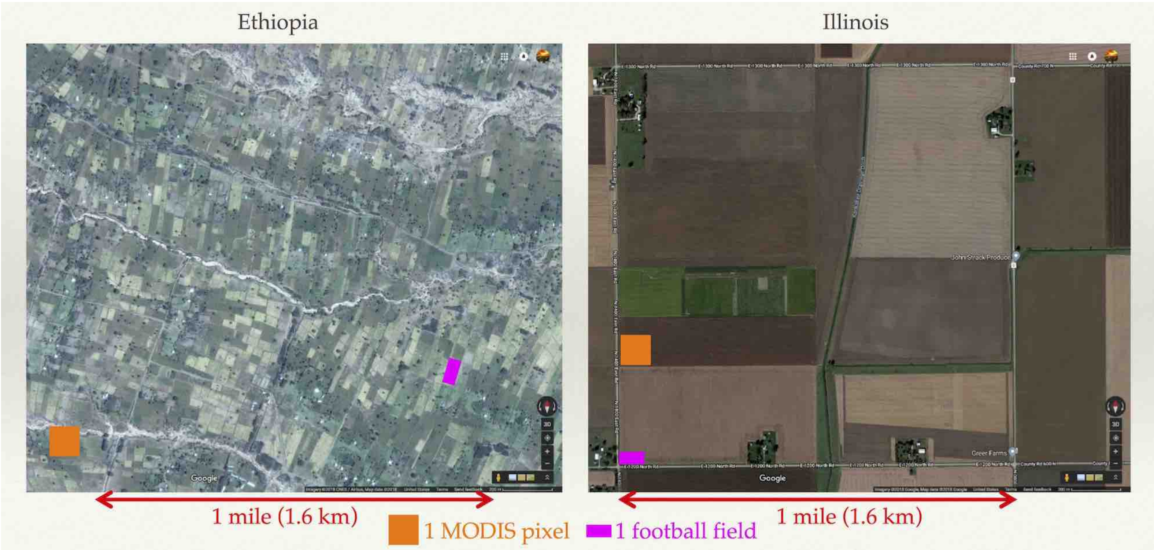

Figure 1. Farm fields by satellite in Ethiopia and Illinois at the same resolution. In Africa, small farm fields (smaller than a MODIS pixel) and poor ground truth data increase the difficulty of analyzing and predicting crop yields. Images are from Google Maps.

Agricultural development is widely known to be crucial for poverty reduction and improved health; thus, there remains a major need to monitor crop health in the developing world [8,9].

Remote sensing has become an asset for detecting environmental changes that impact crop health [10–13]. Today, satellite imagery costs less and is more easily accessible, making remote monitoring more broadly available to scientists and the general public.

However, crop prediction is still very challenging in many African countries due to minimal reporting of crop health and yields; farms consist of very small plots of varied crops interspersed with buildings (Figure 1); and the continent contains a vast number of different climates, growing seasons, and crops [15,16].

A couple of groups currently publish real-time forecasts of crop health. For example, the Group on Earth Observations Global Agricultural Monitoring Initiative (GEOGLAM) [17] and USDA Famine Early Warning System Network (FEWS NET) [18–20] generate advance notice of impending food crises. These systems are comprised of large teams that incorporate data from remote sensing, on-the-ground monitoring, field reports, and agroclimate indicators such as rain, snow, and surface temperatures. These large models require an extensive budget. In contrast to this study, their predictions are simplified into qualitative categories instead of numerical values.

This study differs from previous work in the US and Africa because of its simplicity: it shows that a single user with a personal computer can produce reasonable forecasts of crop yields for the whole continent. I compute an overall measure of relative vegetation health compared to the mean on a per-pixel basis over select subregions in every African country, thus evaluating whether dense farming areas can be used as representative samples of larger regions to increase computational efficiency. Unlike many previous studies, it may be applied anywhere in the world—it does not depend on special tuning for the particular crop, region, or climate of interest. Relatively low-resolution pixels of

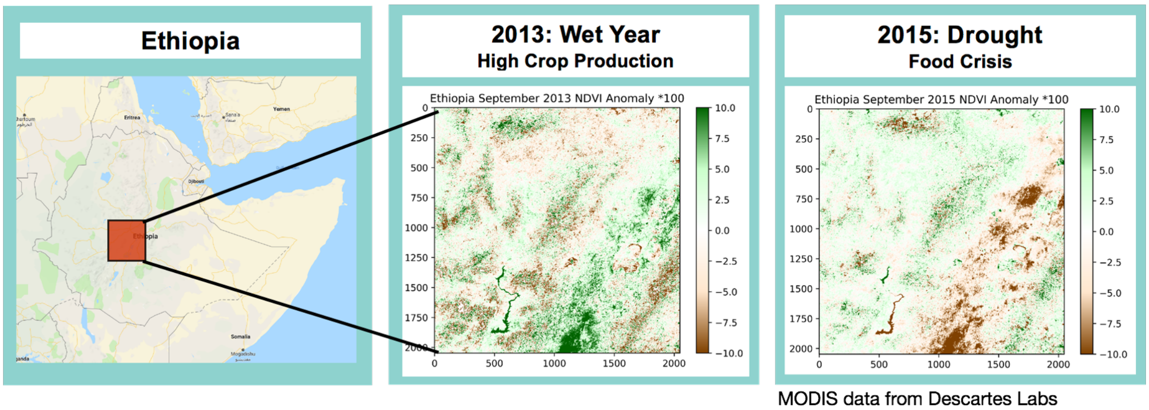

Figure 2. The box examined in Ethiopia and its September NDVI anomalies during a wet year and a dry year. The NDVI anomalies are particularly evident in the Rift Valley, where farming is the most dense.

the Moderate Resolution Imaging Spectroradiometer (MODIS) decrease the amount of data that must be processed, making this system cheaper and more efficient. Crop masks are not used in this model to increase simplicity, versatility, and eliminate the complication of small field sizes, intercropping, and imperfect crop masks. The method was created for developing countries where detailed monitoring on the ground simply does not exist, and was successfully validated against extensive crop data in Illinois.

The goal of this study was to see how well crop yields may be predicted using extremely straightforward methods based on simple averages and differences of common indices over dense farming regions and the resulting correlations. More complex models with crop masks and detailed tuning require a substantial staff and several years to develop and validate. This method, developed and tested by the author over the course of a couple months on a laptop computer, can produce reasonable forecasts of crop yields for the whole continent.

2. Methods

The primary goal of this research is to create a predictive measure of crop yields computed from satellite data. Python code was written to obtain satellite images, mask out clouds, calculate vegetation and water indices (VI), compute monthly VI anomalies since 2000, and correlate the anomalies with crop yield anomalies for every county in Illinois, which served as a proof of concept due to large amounts of ground truth data in the US. The same method was then applied to every country in Africa to create an early indicator of crop yields.

Moderate Resolution Imaging Spectroradiometer (MODIS) imagery was obtained from the Descartes Labs Satellite Platform (Figure 2). I obtained the red, green, blue, and nir infrared (NIR) band for Illinois and Africa for the years 2000-2018. First, clouds were masked out to not be included in the monthly average. To measure the health of crops throughout the growing season, three VIs were computed: NDVI, EVI, and NDWI, defined in Table 1. All three indices have served as crop monitoring tools in previous studies, and have been shown to resemble actual crop conditions. The indices range from -1 to 1, and healthier vegetation registers closer to 1.

Table 1. Definitions of vegetation indices to measure crop health. NIR is near infrared; G is the gain factor; L is the canopy background adjustment that addresses non-linear, differential NIR and red radiant transfer through a canopy; and C_1 and $C2$ are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band.

Index	Description	Formula
NDVI	Normalized Difference Vegetation Index	$\text{NDVI} = \frac{\text{NIR}-\text{Red}}{\text{NIR}+\text{Red}}$
EVI	Enhanced Vegetation Index	$\text{EVI} = G * \frac{\text{NIR}-\text{Red}}{\text{NIR}+C_1*\text{Red}-C2*\text{Blue}+L}$
NDWI	Normalized Difference Water Index	$\text{NDWI} = \frac{\text{Green}-\text{NIR}}{\text{Green}+\text{NIR}}$

For every pixel in Illinois, the VI monthly averages and climatologies were computed. The process begins with daily cloud-masked MODIS swaths. The climatology is defined as the average VI over 2000–2016 for each month and pixel. Next, the monthly climatology is subtracted from the monthly average for every pixel, resulting in the monthly anomaly. The pixels in each county are averaged together to find the county-wide monthly anomaly and county-wide monthly average.

Illinois was chosen as a test site because the land is mostly agricultural and can provide a clear signal of crop health. Illinois also has very little irrigation: most counties irrigate less than 1% of their fields [21]. Similarly, 90% of staple food production in sub-Saharan Africa comes from rain-fed farming systems [22].

Annual crop yield data was downloaded for every county in Illinois for 2000–2016 for three crops: corn, soybeans, and sorghum, from USDA county estimate reports [1]. Because each county has different growing conditions, the mean was subtracted out of each county's crop yield to find the yield anomaly so that comparisons could be made across all counties. Correlations were found between each county's yield anomaly and the three VIs for five months, May–September. To find the highest possible correlation amongst these variables and months, a multivariate regression was fit to each month and index for a total of 15 variables (5 months $\times$ 3 VIs).

To test the predictive ability of the model, the data were split into a training group of 90% and a testing group of the remaining 10%. The multivariate regression was then fit to the training data and asked to predict the testing set. To ensure randomness, this process was repeated ten times for each crop, and the analysis is the composite of these ten prediction sets.

After testing in Illinois was complete, the method was applied to every country in Africa. In each country, the two to four highest-producing crops were analyzed [23]. In each country, a box was analyzed over a dense farming region, which served as a representative sample of the entire country. The VI anomalies and averages from these regions were then correlated to national crop production data [23]. The daily MODIS imagery over the selected boxes in each country was processed in a similar way to Illinois. First, the bands were retrieved from the Descartes Platform. NDVI, EVI, and NDWI were computed, and cloudy pixels were masked out. The climatology for each pixel was subtracted to obtain monthly anomalies as well as averages of all three indices. Next, correlations were computed between the six indices of the month at the height of the growing season and the crop production. The

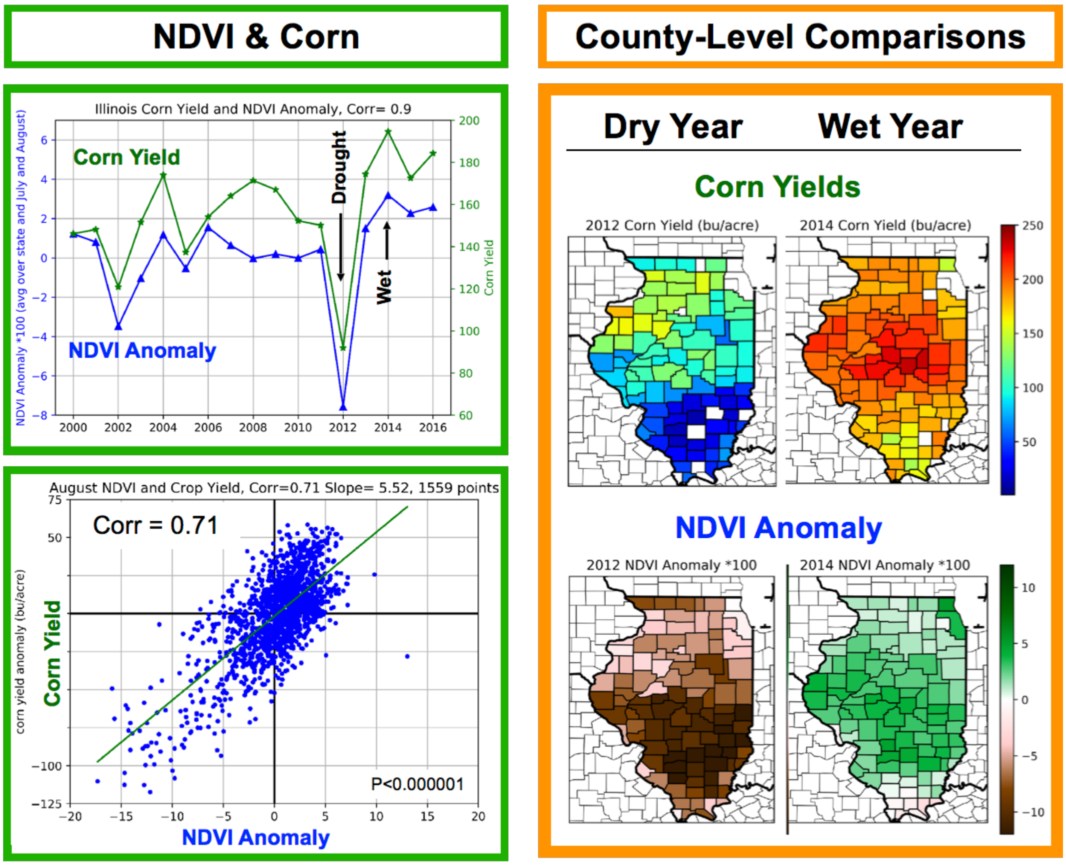

Figure 3. Illinois corn yields versus NDVI anomaly for each year (upper left), as a correlation plot for each county and year (lower left) and for examples of dry and wet years (right).

height of the growing season is defined as the month in the growing season that the NDVI average peaks.

Future crop production for the 2018 harvest was then predicted for every African country. Later, once actual production values were published, the error of the predictions in every country and crop was computed.

3. Results

3.1. Illinois

The method was first validated in Illinois and then applied in Africa. Correlations were computed in Illinois between the anomalies of NDVI, EVI, and NDWI, and three crops: corn, soybeans, and sorghum; all were found to have high correlations. The method was first tested with state-wide averages to show that results are significant when analyzing a large area. The correlations between state-wide corn yield and NDVI, EVI, and NDWI anomalies are extremely statistically significant at 0.90, 0.85, and -0.92 respectively. Figure 3 shows the differences in NDVI anomalies and corn yields during a drought year (2012) and a wet year (2014).

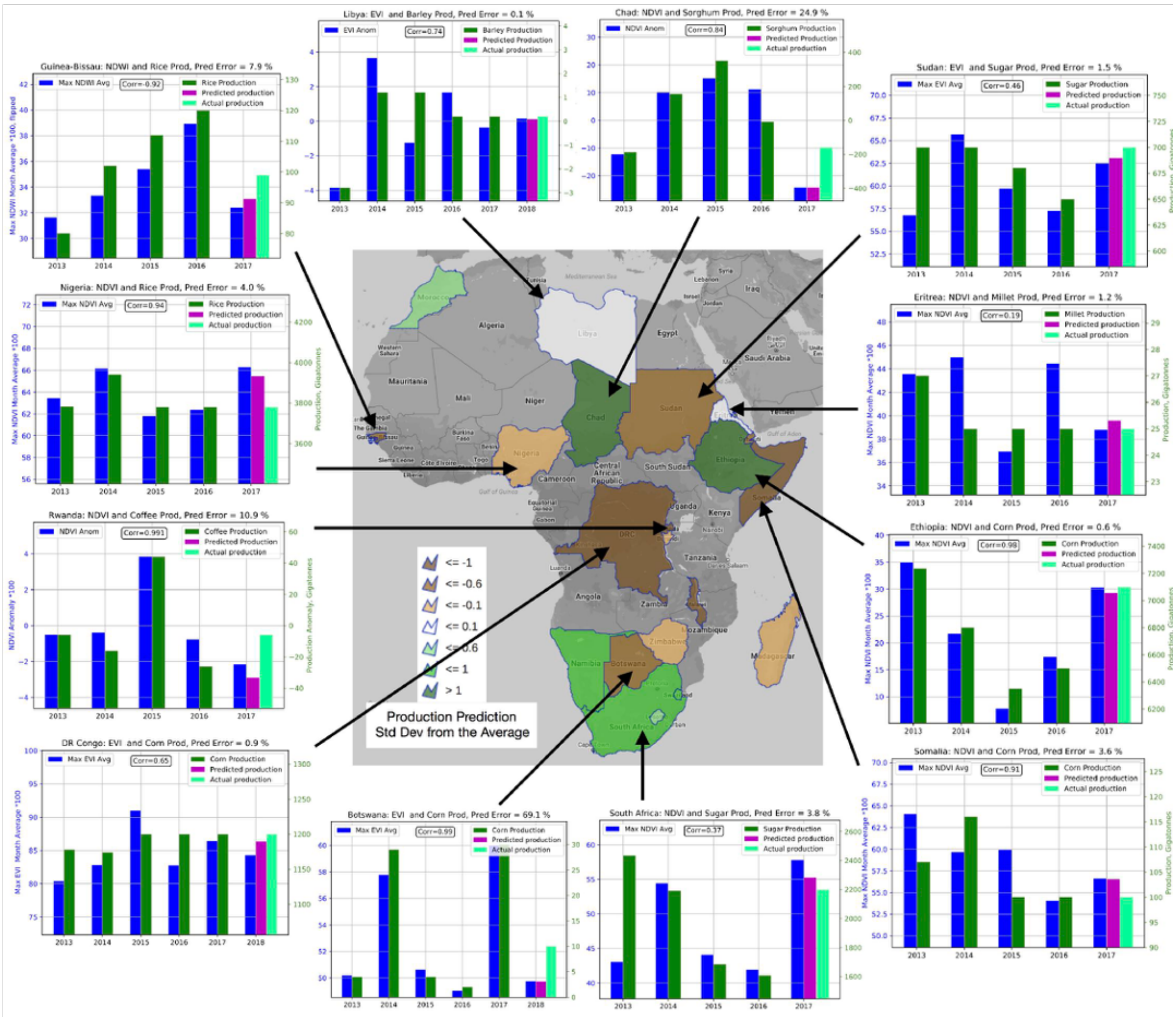

Figure 4. The map in the center displays the predicted crop production in standard deviations from the average. Surrounding the map are bar charts of satellite indices (blue), historical crop production (dark green), predicted 2018 crop production (pink), and actual 2018 production (light green).

To test the predictive power of the model, it was trained on a random 90% of the data and then predicted the remaining 10%. This process was repeated ten times. The model could predict the yield with reasonable error based on only the VI anomalies of the entire county, demonstrating how this simple method is a good indicator of crop yields.

3.2. Africa

The high correlations in Illinois show that this model has good forecasting skill for crop yields. Next, this method was applied to every country in Africa. First, a box in an agricultural region was selected in each country and a total of 10 terabytes of daily satellite imagery was processed according to the method above. Correlations and linear regressions were computed in every country for their 2–4 highest producing crops. Difficulties in finding accurate correlations include:

- false reporting of production in some countries due to lack of resources, poor oversight, or corruption (e.g., DR Congo, Eritrea, Libya). In severe cases, one could simply use the NDVI anomaly as a proxy for production rather than computing a correlation with reported crop yields.
- multiple growing seasons in specific central countries (Rwanda, Somalia);
- poor quality of earth observation data (e.g., clouds) every day for months at a time in central African countries (Gabon, Cameroon) [14]; and
- time delays and misclassification of harvests during October–December, where production is incorrectly reported in the following calendar year (Nigeria, Sudan).

In each African country, correlations were computed between every crop and six indices (NDVI, EVI, NDWI, monthly averages and anomalies). Next, the historical regressions were used to predict crop production for 2018 harvests. Every country that reported productions for their 2018 harvest before the publication of this article was examined. This includes harvests ranging from December 2017 (e.g., Ethiopia) through June 2018 (e.g., Namibia), and included a total of 21 countries, about half of Africa. In April 2018, VI anomalies and crop predictions were posted on a publicly viewable interactive map [24], and the actual production values were added as they became available in mid to late 2018 (Figure 4).

In Ethiopia, the model predicted the 2018 harvests to yield 7055 giga-tonnes (GT) of corn and 4174 GT of sorghum. The actual production was 7100 GT and 4100 GT, respectively, for an error of 0.6% and 1.8%. These minimal errors show how this simple model can predict yields very accurately, even with only a few years of historical relationships.

Small errors in predictions were common across Africa. The median error was 8.6%. Twenty-one percent of the predictions had a relative error below 2%, and 40% had errors below 5%.

4. Conclusions

In this research, I developed a method to predict crop yields 2–4 months before the harvest, based on daily MODIS satellite imagery. The model was first validated in Illinois where there is county-level crop yield data by computing the linear fit between yields and VIs. When a split-sample validation was applied to a multivariate regression with all months of the growing season and all three VIs, the model could predict the crop yields within 5.7%, 5.8%, and 22% for corn, soybeans, and sorghum, respectively. After this success, satellite imagery was analyzed in every African country, and productions for the 2018 harvests were predicted 2–4 months before the harvest. Once 2018 harvests were published, the prediction accuracy was tested against the reported values. Forty percent of the predictions were found to have less than a 5% error.

The main objective of this study was to show how a very simple method can serve as an early warning system to predict crop yields in every African country. This method relies solely on NDVI, EVI, and NDWI anomalies calculated over specific subsections of the countries, without the use of crop masks, subnational yield statistics, or special tuning for location or climate. Even with these many simplifications, the model was still able to produce predictions with reasonable error over Illinois and throughout Africa.

The prediction accuracy for different crops varies substantially. Some crops are harder to predict, as each crop correlates to the VIs with different strengths. Some crops may also be affected by extreme weather late in the season, which this model does not include since it predicts yields from the height during the growing season. Millet, sugar, and rice had the lowest errors, while cotton, wheat, and sorghum were much harder to predict.

A limitation of this model is that it relies on published yield data, so it will not predict as reliably in countries that lack reporting accuracy. In these places, the NDVI anomaly could be used as a proxy for relative crop yields compared to a mean. The model also only predicts yields at the national level and has no subnational component. However, it has the ability to predict yields sub-nationally in the future when sub-national crop data are supplied.

In this study, country-wide crop production was correlated to VI anomalies over dense farming regions to test if small areas could serve as representative samples of the entire country. In most countries, the subregions only covered between 1% and 15% of the total land, depending on the size of the country and box. Despite these small areas, the model produced surprisingly high correlations between the VIs and crop production. South Africa is an exception, with low correlations and high errors in the predictions. South Africa has farms across the country, so the selected box was not able to represent the entire area. In many other African countries, one region is a primary producer and can be used to predict country-wide production.

The model developed here may be compared to the existing early-warning systems of GEOGLAM and FEWS NET. Both are run under large budgets by an extensive team of people with partnerships around the globe. Their systems include local surveyors, remotely sensed data, agroclimate indicators, field reports, and communications with national and regional experts. In contrast, this method can be run by a single user on a modern laptop computer. It was developed over the course of a couple months, and is practically free. This model is also able to predict a numerical value of crop production, while GEOGLAM and FEWS NET present their results as a qualitative measure: conditions are compacted into five categories of crop conditions or food insecurity phases.

The power of the method developed here is that it can be applied to any crop, location, or climate to produce reasonable real-time forecasts of crop yields. It is unique because of its versatility and easy to apply due to its simplicity.

References

1. Hamer, H.; Picanso, R.; Prusacki, J.J.; Rater, B.; Johnson, J.; Barnes, K.; Parsons, J.; Young, D.L. USDA/NASS QuickStats US Crop Data. 2017. Available online: https://quickstats.nass.usda.gov (accessed on 30 October 2018).
2. Menne, M.J.; Durre, I.; Vose, R.S.; Gleason, B.E.; Houston, T.G. An Overview of the Global Historical Climatology Network-Daily Database. *J. Atmos. Ocean. Technol.* **2012**, *29*, 897–910, doi:10.1175/JTECH-D-11-00103.1.

3. Petersen, L.K. America's Farming Future: The Impact of Climate Change on Crop Yields. *AMS* **2018**. doi:10.5281/zenodo.1473201.
4. McKinnon, K. GHCN-D: Global Historical climatology Network daily temperatures NCAR—Climate Data Guide. 2016. Available online: https://climatedataguide.ucar.edu/climate-data/ghcn-d-global-historical-climatology-network-daily-temperatures (accessed on 30 October 2018).
5. Carletto, G.; Beegle, K.; Himelein, K.; Kilic, T.; Murray, S.; Oseni, M.; Scott, K.; Steele, D. Improving the Availability, Quality and Policy-Relevance of Agricultural Data: The Living Standards Measurement Study Integrated Surveys on Agriculture. In Proceedings of the Third Wye City Group Global Conference on Agricultural and Rural Household Statistic, Washington, DC, USA, 8–9 April 2008; p. 26.
6. Carletto, C.; Jolliffe, D.; Banerjee, R. From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *J. Dev. Stud.* **2015**, *51*, 133–148, doi:10.1080/00220388.2014.968140.
7. Challinor, A.; Wheeler, T.; Garforth, C.; Craufurd, P.; Kassam, A. Assessing the vulnerability of food crop systems in Africa to climate change. *Clim. Chang.* **2007**, *83*, 381–399, doi:10.1007/s10584-007-9249-0.
8. Conceicao, P.; Levine, S.; Lipton, M.; Warren-Rodriguez, A. Toward a food secure future: Ensuring food security for sustainable human development in Sub-Saharan Africa. *Food Policy* **2016**, *60*, 1–9, doi:10.1016/j.foodpol.2016.02.003.
9. The World Bank. *World Development Report 2008: Agriculture for Development*; The World Bank: Washington, DC, USA, 2008.
10. Maas, S.J. Using Satellite Data to Improve Model Estimates of Crop Yield. *Agron. J.* **1988**, *80*, 655–662, doi:10.2134/agronj1988.00021962008000040021x.
11. Hellden, U.; Eklundh, L. *National Drought Impact Monitoring—a NOAA NDVI and Precipitation Data Study of Ethiopia*; Technical Report; Lund University Press: Lund, Sweden, 1988.
12. Gao, B.C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, *58*, 257–266, doi:10.1016/S0034-4257(96)00067-3.
13. Lobell, D.B. The use of satellite data for crop yield gap analysis. *Field Crops Res.* **2013**, *143*, 56–64, doi:10.1016/j.fcr.2012.08.008.
14. Atzberger, C.; Rembold, F. Mapping the Spatial Distribution of Winter Crops at Sub-Pixel Level Using AVHRR NDVI Time Series and Neural Nets. *Remote Sens.* **2013**, *5*, 1335–1354, doi:10.3390/rs5031335.
15. Gissila, T.; Black, E.; Grimes, D.I.F.; Slingo, J.M. Seasonal forecasting of the Ethiopian summer rains. *Int. J. Climatol.* **2004**, *24*, 1345–1358, doi:10.1002/joc.1078.
16. Tadesse, T.; Demisse, G.B.; Zaitchik, B.; Dinku, T. Satellite-based hybrid drought monitoring tool for prediction of vegetation condition in Eastern Africa: A case study for Ethiopia. *Water Resour. Res.* **2014**, *50*, 2176–2190, doi:10.1002/2013WR014281.
17. Becker-Reshef, I.; Justice, C.; Sullivan, M.; Vermote, E.; Tucker, C.; Anyamba, A.; Small, J.; Pak, E.; Masuoka, E.; Schmaltz, J.; et al. Monitoring Global Croplands with Coarse Resolution Earth Observations: The Global Agriculture Monitoring (GLAM) Project. *Remote Sens.* **2010**, *2*, 1589–1609, doi:10.3390/rs2061589.
18. Senay, G.; Velpuri, N.; Bohms, S.; Budde, M.; Young, C.; Rowland, J.; Verdin, J. Chapter 9—Drought Monitoring and Assessment: Remote Sensing and Modeling Approaches for the Famine Early Warning Systems Network. In *Hydro-Meteorological Hazards, Risks and Disasters*; Shroder, J.F., Paron, P., Baldassarre, G.D., Eds.; Elsevier: Boston, MA, USA, 2015; pp. 233–262,
19. Funk, C.; Verdin, J.P. Real-Time Decision Support Systems: The Famine Early Warning System Network. In *Satellite Rainfall Applications for Surface Hydrology*; Springer: Dordrecht, The Netherlands, 2010; pp. 295–320,
20. Molly, E.; Brown, E.B.B. Evaluating the use of remote sensing data in the U.S. Agency for International Development Famine Early Warning Systems Network. *J. Appl. Remote Sens.* **2012**, *6*, doi:10.1117/1.JRS.6.063511.
21. United States Department of Agriculture, National Agricultural Statistics Service. *Farms and Farmland: Numbers, Acreage, Ownership, and Use*; United States Department of Agriculture, National Agricultural Statistics Service: Washington, DC, USA, 2014.
22. Cooper, P.J.M.; Dimes, J.; Rao, K.P.C.; Shapiro, B.; Shiferaw, B.; Twomlow, S. Coping better with current climatic variability in the rain-fed farming systems of sub-Saharan Africa: An essential first step in adapting to future climate change? *Agric. Ecosyst. Environ.* **2008**, *126*, 24–35, doi:10.1016/j.agee.2008.01.007.
23. Mundi, I. Agricultural Production Statistics by Country. 2018. Available online: https://www.indexmundi.com/agriculture (accessed on 30 October 2018).

24. Petersen, L.K. Predicting Food Shortages in Africa from Satellite Imagery. 2018. Available online: https://lillianpetersen.github.io/africa_satellite (accessed on 30 October 2018).