CHAPTER 9

Remote Sensing and Machine Learning for Food Crop Production Data in Africa Post-COVID-19

Racine Ly, Khadim Dia, and Mariam A. Diallo

Introduction

The world is experiencing an unprecedented health crisis during the spread of COVID-19 (SARS-CoV-2, or Severe Acute Respiratory Syndrome Coronavirus 2). While the pandemic appears to be less severe on the African continent than in other geographic regions[1] (Global Change Data Lab 2021), its economic impact is significantly more pronounced. COVID-19 is upending livelihoods, damaging business and government balance sheets, and threatening to reverse development gains and growth prospects for years to come in Africa south of the Sahara (IFC 2020). The World Bank forecasts that Africa south of the Sahara will go into recession in 2020 and that COVID-19 will cost the region between $37 billion and $79 billion in output losses in 2020 alone. The informal sector, a significant source of income and employment, will be the hardest hit.

In the agricultural sector, COVID-19 threatened to lead to a severe food security crisis in the region in 2020, with disruptions in the food supply chain and expected contractions of 2.6 to 7 percent in agricultural production (Zeufack et al. 2020). Travel bans, border closures, and the late reception and use of agricultural inputs such as imported seeds, fertilizers, and pesticides could lead to poor performance in food crop production. Another layer of disruption introduced by the mobility restriction measures is the scarcity of agricultural workers, mainly seasonal workers. Lockdown measures and border closures limit seasonal workers' ability to get to farms in time for planting and harvesting activities (Ayanlade and Radeny 2020; ILO 2020). Moreover, delivery of most of the imported agricultural inputs relies on air travel, which has been impacted heavily by the pandemic (Vilardell and Baenas 2020). Such transportation disruptions can also negatively affect the food crop production system.

It is challenging to fully understand the relationships between the COVID-19 containment measures taken by countries and their impacts on food crop production. Comprehending these relationships would require studies on the impacts of the containment measures on farmers' and seasonal workers' mobilities and the prompt reception of seeds, fertilizers, and pesticides for cropping activities. The kinds of datasets that would allow these studies are not yet available to the best of our knowledge. However, it is risky to wait to take action until food crop production statistics are available at the end of the agricultural season. Instead, it would be better to have an estimate of the most likely food crop production levels before the harvesting period to allow for better planning and early policy actions. For that goal, data are most needed.

Access to reliable and timely data in the agricultural sector has been problematic in Africa for a long time. Even in regular times, there are difficulties in accessing agricultural statistics. The issue is even more pronounced in crisis times, such as the current pandemic, when, paradoxically, the data are most needed. Uninformed decision-making is the most significant consequence of the lack of data and analytics. Making decisions based on anecdotal facts creates inefficiencies in problem-solving. Much of the assumed knowledge about agriculture in Africa may no longer be valid, given Africa's rapid economic transformation, fast urbanization, demographic and climatic changes, and, more importantly, the scarcity of quality data (Christiaensen and Demery 2018). In a rapidly changing world, the facts that drive research and policy focus quickly become outdated. COVID-19 highlighted the need to improve African food systems' resilience. Access to timely, spatially disaggregated, and accurate agricultural statistics can play a significant role in achieving that goal. That is the main focus of this chapter.

This chapter assesses food crop production levels in 2020—before the harvesting period—in all African regions and for staples such as maize, cassava, rice, and wheat. Production levels are predicted using the combination of remote sensing data retrieved from satellite images and a machine learning artificial neural networks (ANNs) technique. The remote-sensing products are used as input variables in the Africa Crop Production (AfCP) model developed by AKADEMIYA2063. The input variables are the normalized difference vegetation index (NDVI), the daytime land surface temperature (LST-day), rainfall data, and agricultural lands' evapotranspiration (ET). The model's outputs are pixel-level maps of agricultural production forecasts for major crops in all African countries. The product and input time-series data are made publicly available on a web-based platform, the Africa Agriculture Watch, to facilitate access to such information for policymakers and other stakeholders.

The chapter is organized as follows: Section 2 provides the underlying conceptual framework that explains the basis for the use of remote sensing products (RSPs) and machine learning for resilient food systems. Section 3 introduces all variables that have been considered for the predictive model, the methodology used to select the crops for each region, and a methodological

1 As of April 19, 2021

description of the machine learning predictive modeling framework. Section 4 presents the predicted food crop production for each region and crop and discusses potential factors related to COVID-19 that might have caused a decline in food crop production in some African countries. Section 5 outlines the limitations of the model and the direction of future research. Finally, Section 6 proposes recommendations to strengthen African food system resilience through an improved data environment and analytics using emerging technologies.

The Basis for the Use of RSPs and Machine Learning for Resilient Food Systems

A resilient food system is determined by its capacity to withstand and recover from disruptions and to ensure a sufficient food supply for communities. Another aspect of food system resilience is the availability of evidence-based technical assistance to help policymakers and decision-makers more effectively prepare for and respond to shocks. Technological advancements can help with that goal. Remotely sensed data via satellite images are now democratized and show a sufficiently high spatial resolution to include a large proportion of agricultural lands across the continent, and machine learning techniques offer a way to build robust predictive models relieved from rule-based approaches. This section provides a conceptual framework for understanding the building blocks of our approach to using RSPs and predictive modeling through machine learning techniques for better-informed policymaking in a time of crisis such as the COVID-19 pandemic.

The Basis for the Use of RSPs for Decision-Making in Agriculture

Real-time (or near real-time) data gathering and analysis is crucial to providing a clear picture of any crisis dynamic and monitoring the effects of simultaneous shocks. The availability of accurate and frequently updated data that reflect the status on the ground requires significant coordination and collaboration, and robust data systems.

In the African context, the use of remote sensing in the agricultural sector is hindered by a lack of reliable ground-truth data. The cost of generating ground-truth data is usually prohibitive, primarily due to the size of the continent. Moreover—and most importantly—the African food production system is characterized by scattered, small-size farms with significant crop spatial heterogeneity. For instance, most African farmers are smallholders with farm sizes of fewer than 2 hectares on which several crops are grown simultaneously. Such a complex cropping system does not facilitate ground truthing for machine learning applications. The diversity of agroecological zones adds another layer of complexity, as does the frequent cloud cover. These challenges often do not allow satellites' optical sensors to collect ground spectral signatures for an extended period of time in some countries.

The levels of data quality, frequency, and disaggregation do not allow a thorough analysis of cropping activities, early anomaly detection, and forecasting for African food production systems. Through Earth observation, RSPs show promise in significantly reducing the underlying gaps in data quality, size, disaggregation, and frequency. RSPs are used in two main ways in agricultural policymaking. First, they are used to provide disaggregated views of agricultural lands and their corresponding biogeophysical parameters. Second, they are used to monitor the effects of agricultural policies on the ground.

RSPs use the spectral signature of earth's features to monitor agricultural lands. Vegetation indexes, such as the NDVI derived from satellite images, provide an assessment of the health of crops by using measurements of the rate of leaves' infrared reflection as a proxy for their visible light absorption rate (light absorption is needed for photosynthesis). RSPs measure several other biogeophysical parameters related to food crop production, such as ET (Running et al. 2021) and LST-day (Wan et al. 2015), and provide data for indexes like the enhanced vegetation index (EVI) (Didan 2015). In general, the combination of measurements of specific spectral layers allows a determination of the agricultural land's biogeophysical status at a community level (see Figures 9A.1, 9A.2, and 9A.3 in the appendix for NDVI, LST-day, and rainfall anomalies for 2020).

Using RSPs to monitor changes on the ground due to agricultural policy has been successful in other parts of the world. Harnessing moderate resolution imaging spectroradiometer (MODIS) NDVI time-series signals, Lein (2012) showed how a tax-free agricultural ordinance in 2006 impacted the adoption of multiple cropping practices in China. Arvor and colleagues (2011) derived indexes from satellite images to study the relationship between agricultural dynamics in Amazonia and the region's existing public policies during the period from 2000 to 2007.

Moreover, satellites revisit the same area many times a year—typically, every five days for Sentinel-2[2] and daily for MODIS[3] —allowing practitioners to monitor land-use and land-cover changes (Li et al. 2020), identify crop taxonomy (Kpienbaareh et al. 2021) and cropping activities (Rezaei et al. 2020), and assess surface water availability (Pekel et al. 2016). Another added value of using RSPs to improve agricultural statistics is the capacity to provide disaggregated information at a pixel level and disentangle the data from conventional administrative entity-based maps. Several weeks (or months) would be necessary to cover the same area with field agents, and still, the results would be less accurate. This capacity to provide community-level detail on maps could allow for targeted responses where they are needed the most. However, RSPs alone cannot provide estimates of potential future agricultural production and yield—that requires a predictive modeling framework.

The Basis for the Use of Machine Learning for Decision-Making

General-purpose technologies have triggered a wide range of innovations globally. The fast pace of technological advances has reduced the cost of technology products and services, encouraged wide adoption, and significantly increased data generation over the last three decades. Combined with advances in computer modeling, these advances have opened up a new "technium" (that is, the accumulation of inventions that humans have created, and which society depends on as much as nature) of data-driven technologies and machine learning techniques.

Machine learning is a set of techniques particularly suitable for making predictions under certain circumstances. These techniques have the capacity to mimic key characteristics attributed to human intelligence, such as vision, speech, and problem-solving. Several papers have shown how machine learning models outperform humans in accuracy in some tasks (Buetti-Dinh et al. 2019; Mnih et al. 2013; Silver et al. 2016). This performance has been possible due to the combination of significant increases in data availability, improvements in computational power, and advancements in algorithmic techniques in the last three decades. The most commonly used supervised-learning technique uses examples and experience to teach how humans make predictions. The old approach of transferring human knowledge to machines through sequential steps is being replaced with an approach of providing machines with data—the same data humans have access to and use to learn. Since a machine does not have to learn other tasks (as a human does), its resources are fundamentally oriented toward learning the relationship between the input data and the corresponding outcomes. The result is a faster learning process and better accuracy in a specific task.

As in previous technological revolutions, the most significant impact would be expected in sectors that are not traditional users of these technologies, such as agriculture. Machine learning techniques can support efforts to forecast agricultural productions and yields (Ly and Dia 2020; Kaneko et al. 2019), manage natural resources, and reduce uncertainty and risk across the agricultural sector. African farmers are mostly smallholders (Conway, Badiane, and Glatzel 2019) facing significant uncertainties that can lead to poor performance, such as erratic rainfall, lack of knowledge about biogeophysical parameters and soil water content, and inadequate planting periods. The capacity to forecast agricultural production given these uncertainties is pivotal for farmers, planners, and policymakers.

Prediction is at the heart of decision-making; however, predictions are just one component of the process. The other decision-making components are judgment, action, outcome, and three types of data that include input, training, and feedback (Agrawal, Gans, and Goldfarb 2018). When decision-makers have access to the same input and training datasets and the same feedback loop, the two key factors that impact their interventions are judgment and predictions based on context. While judgment is a subjective concept that depends on background and experience, predictions can be objective and follow mathematical formulations; therefore, they can be improved faster and enhance the entire process of designing and implementing informed strategies.

2 Sentinel-2 is an Earth observation mission from the Copernicus Programme that systematically acquires optical imagery at high spatial resolution (10 m to 60 m) over land and coastal waters. The mission is currently a constellation with two satellites, Sentinel-2A and Sentinel-2B, operated by the European Space Agency.

3 MODIS is the key sensing instrument aboard the Terra (originally known as EOS AM-1) and Aqua (originally known as EOS PM-1) satellites. Terra MODIS and Aqua MODIS are capturing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups of wavelengths.

Combining RSPs and Machine Learning for Resilient Food Systems

The path from RSPs and machine learning to policymaking is not straightforward, especially in Africa. The requirements for sustainable use of RSPs in policymaking for the agricultural sector (or any sector that requires RSPs) necessitate political will, technical expertise in remote sensing and data analysis, and the institutional strength, including sufficient financial and infrastructural resources, to deal with those tasks. Our rationale for combining RSPs and machine learning to build resilient food systems is as follows: the spatial and temporal resolutions of RSPs allow a disaggregated view of agricultural lands, with several indicators that assess crop growing conditions at a community level. As inputs into the machine learning predictive modeling framework, RSPs are expected to contribute to the development of reasonably accurate predictions about food crop production before the harvesting period. This helps build a more resilient food system by improving knowledge about potential agricultural production at the community level.

Public availability of consolidated, ready-to-use biogeophysical RSPs and food crop production forecast maps would reduce the technical, infrastructural, and institutional barriers that have the potential to prevent African countries from using RSPs and machine learning for resilient food systems. Table 9.1 shows the underlying techniques and concepts that must be harnessed to provide near real-time biogeophysical data and food crop production maps at the community level for all African countries. The corresponding outputs, outcomes, and constraints involved in decision-making in food production systems are illustrated.

TABLE 9.1—TECHNIQUES, OUTPUTS, AND OUTCOMES TO HARNESS RSPS AND MACHINE LEARNING FOR DECISION- AND POLICYMAKING FOR FOOD CROP PRODUCTION

Technique and Concept	Output	Outcome and Constraints Lifted
The use of RSPs and machine learning to assess policies impacts on food crop production systems.	Use the time-series data provided on the web-based tool to assess if the policy goals are reached or not, and take corrective actions.	
Decision-making and policy making based on forecasts and biogeophysical parameters time series.	Use the food crop production disaggregated forecast map at the community level to plan and strategize based on the scenario provided by the model.	All the information based on RSPs and their most likely future outcomes are made available to decision-makers to provide impactful policies.
The use of web-based tools to make datasets and maps publicly available.	Make the food crop production forecast maps and time-series data used as inputs available in a web-based tool.	Remote sensing products and food crop production forecasts at the community level are made publicly available, lifting the data access constraint.
Forecasts based on the combination of RSPs and ML and third-party data.	Use the combination of input variables data and ML method to learn the data patterns and use the data structure learned for future predictions.	The technical field-level expertise needed for a machine learning predictive model is not a constraint anymore.
Remote sensing products from satellite images and machine learning techniques.	Make available preprocessed input maps such as NDVI, LST, rainfall, ET, production maps as labels, and crop masks and choose ML technique.	The lack of technical skills related to data processing methods for satellite image is not a constraint anymore for analysts, decision-makers and policymakers.

Source: Authors.
Note: ET = evapotranspiration; LST = land surface temperature; ML = machine learning; NDVI = normalized difference vegetation index; RSPs = remote sensing products.

Remotely Sensed Data, Crop Selection, and Predictive Modeling Framework

Food crop production estimations based on remote sensing can be built through two main approaches: (1) using remotely sensed data as inputs into agro-meteorological or plant-physiological models, and (2) building a direct mathematical relationship between remotely sensed data and crop production (Huang and Han 2014). The first approach is based on mechanistic descriptions of crop growth, development, and production simulated through mathematical functions. Methods like this have shown satisfactory results but cannot exploit datasets to their full extent due to constraints related to the way crop growth phenomena are described with mathematical functions. The second approach usually relies on derived indicators from remotely sensed data and their correlation with crop growth and yield. In the analysis in this chapter, we follow the second approach to predict food crop production values based on inputs derived from RSPs. In this section, we describe the data used, the selection of crops considered, and the construction of the predictive modeling framework, AfCP.

Biogeophysical Remotely Sensed Data for Food Crop Production Forecasts

One of the most known and used parameters to characterize vegetation cover is the NDVI, derived from near-infrared and red bands from multispectral sensors. The NDVI indicator is extensively used to characterize vegetation cover due to its close relationship with several vegetation parameters such as leaf area index, the fraction of absorbed photosynthetically active radiation, and green biomass. Many studies have been conducted to predict crop yield from NDVI signals (see Liu et al. 2019; Rembold et al. 2013; Rasmussen 1992, 1997). However, there are limitations to using only NDVI as a proxy for crop yield estimation due to its dependencies on the crop, soil, and leaf types. Indeed, even though NDVI is a good proxy for aboveground biomass production, the relationship between biomass and yield varies in time and space (Leroux et al. 2016). Our approach emphasizes the use of several RSPs, and thus ensures the use of even more information about crop status than the use of only NDVI.

Several studies conducted in the 1970s have shown that final crop yield can be related to thermal indexes (Idso, Jackson, and Reginato 1977; Smith et al. 1985). Because of this, an LST-day layer has been used as a proxy for crop water stress in our methodology. Water availability is also a key component for crop growth and yield; therefore, it is essential when building a crop production model to take it into account. However, in most African countries, agricultural lands are rainfed (Stockholm International Water Institute 2018), so rainfall data has been derived from the climate hazards group infrared precipitation with station data RSPs.

Soil water content and its dynamic in the ground under agricultural lands is an important parameter to address. The underground water is conveyed toward the atmosphere through two main channels: evaporation and transpiration. The former corresponds to the transformation of liquid water into a gaseous state and its release into the atmosphere. For the evaporation process to occur, soil moisture, vapor pressure gradient, and 600 calories of heat energy for every 1 gram of water are required. The transpiration mechanism consists of underground water transportation from the soil to a plant's roots, then from the roots to the leaves through the vascular plant tissues, and ultimately, from plant tissues to evaporation into the atmosphere. As explained by Bhatt and Hossain (2019), transpiration is the most desired mechanism since water transportation through the plant's internal structure also carries nutrients from the soil to the plant and prevents the plant's tissue from overheating. However, measuring the two processes' contributions to the amount of water in the atmosphere is difficult; therefore, their combined effects are usually measured with the ET index from RSPs. We use the total ET of crop locations as a proxy for measuring the crops' rate of transpiration, which, by its effectiveness, will inform us of the crops' health.

Our production-estimation methodology allows us to predict production for one crop in the region of interest before the harvesting period and at the community level. Raster-type maps for historical production quantities for 42 crops and at a global scale are publicly available from the spatial production allocation model (SPAM) database (IFPRI 2016, 2019, 2020). They have been generated by an allocation model with a grid cell size of 10 kilometers. These maps are used for two purposes: (1) using the pixel production values as response variables to our model and (2) creating crop masks to target areas where a specific crop is believed to be grown. Table 9.2 summarizes the list of RSPs taken as inputs and

TABLE 9.2—INPUT PARAMETERS FOR THE FOOD CROP PRODUCTION MODEL WITH THEIR SPATIAL AND TEMPORAL CHARACTERISTICS

Input Parameters	Dataset ID	Spatial Resolution (km)	Temporal Resolution (days)	Temporal Extent (period)
NDVI	MOD13A2	1.00	16	2000–NOW
LST-day	MOD11A2	1.00	8	2000–NOW
Rainfall	Africa_monthly	5.55	30	1981–DEC 2020
ET	MOD16A2	0.50	8	2000–NOW
Production	P	10.00	—	2000, 2005, 2010, 2017

Source: Authors.

response variables for our food crop production model and their spatial and temporal characteristics.

Crop Selection for the Food Crop Production Model

African farmers are mostly smallholders who grow food for consumption and income. Because of chronic infrastructural and financial issues and difficulties in accessing agricultural inputs and markets, a relatively low-intensity shock could significantly impact their food security status. Therefore, having knowledge of potential future agricultural production before the harvesting period is essential for planning purposes. Crops should be targeted according to their relative importance for a country's most vulnerable communities, especially during a major crisis such as the COVID-19 pandemic.

In this chapter, the criteria that determined the importance of a food crop were a combination of its production quantities and the food self-sufficiency of a country. Two rankings were performed to identify a country's top 5 important food crops: the first was a ranking of the 10 most-produced food crops, and the second was a ranking of food crop self-sufficiency. The analysis relied on agricultural variables that were publicly available on international databases. Production and domestic supply data for 2014–2018 (the most recently available years) were available from the Food and Agriculture Organization (FAO 2021).

For each country, a preliminary list was developed of the 10 most-produced

TABLE 9.3—SELECTED LIST OF FOOD CROPS BY AFRICAN REGION

African Regions (# of Countries)	Food Crops	# of Countries where Crop Is in the Top 5
Eastern Africa (14)	Maize	8
	Cassava	8
	Sugarcane	9
Central Africa (7)	Cassava	5
Southern Africa (5)	Maize	3
Northern Africa (5)	Wheat	4
Western Africa (16)	Cassava	8
	Rice	9
	Maize	7

Source: Authors.

agricultural commodities in terms of quantities. Subsequently, the sufficiency aspect was included through the self-sufficiency ratio, defined by the share of food crop consumption to food crop production at the domestic level. The ratio evaluated whether a country produced enough food crops to cover its own needs for each of the 10 most-produced food crops. An agricultural commodity was considered essential for a country if the consumption was greater than the production. The list of the 5 primary produced and consumed commodities for each country by region are reported in Tables 9A.1, 9A.2, 9A.3, 9A.4, and 9A.5 in the appendix.

The most predominant in the top five food crops among countries were selected for the regional level. Table 9.3 presents the selected list of crops for each region. In eastern Africa, maize, cassava, and sugarcane were selected as the major food crops. Indeed, 8 of 14 countries, based on our ranking, had maize and cassava as their top priority products in terms of production and consumption. Sugarcane was also essential for 9 of 14 eastern African countries. In western Africa, three crops were selected: cassava, rice, and maize. Cassava, maize, and wheat were identified as essential in central Africa, southern Africa, and northern Africa, respectively. Due to data representativeness considerations, only maize was considered for eastern Africa.

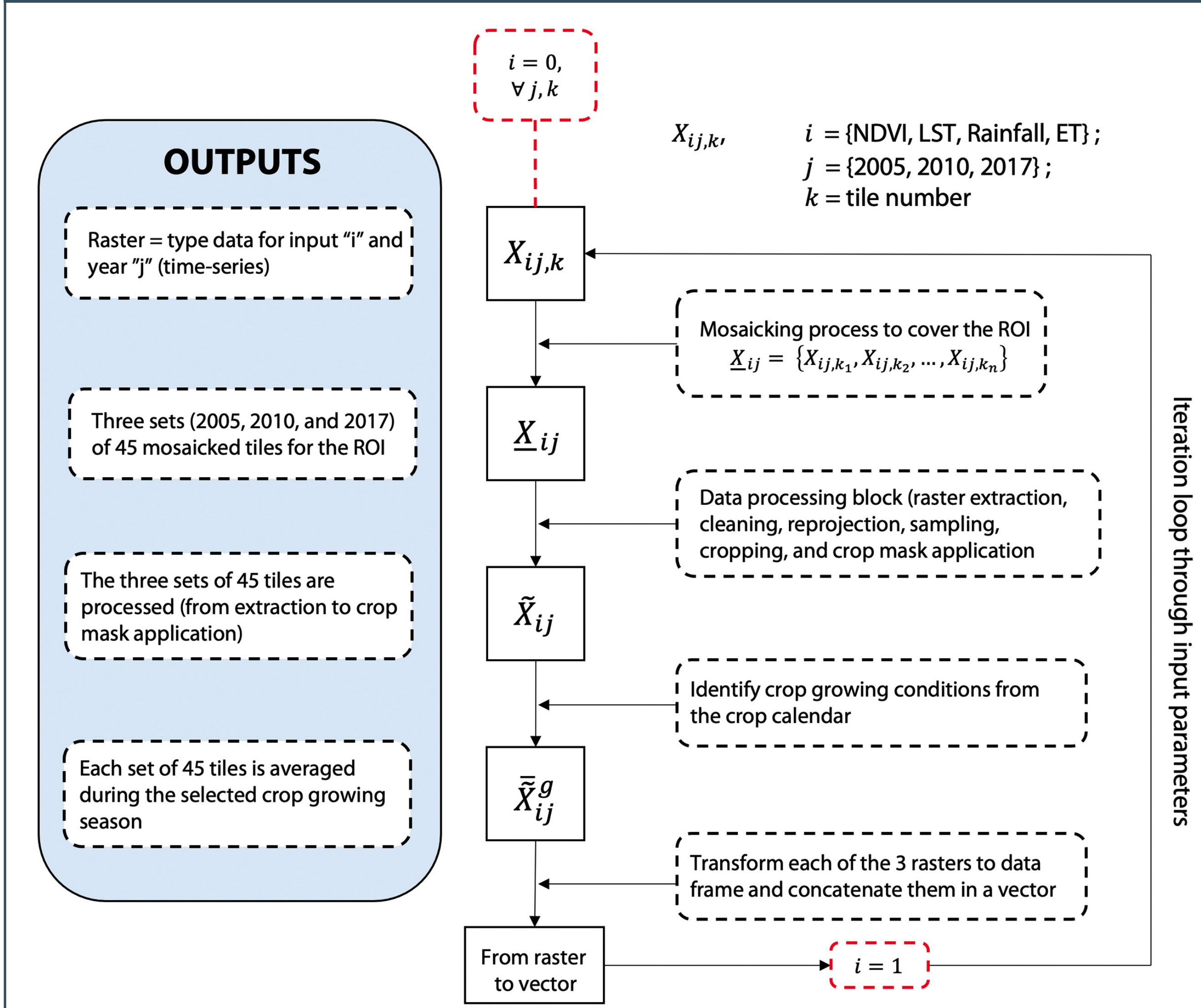
Predictive Modeling Framework

Data Preprocessing and Input Variable Prediction

Data preprocessing procedures were carried out to build the final dataset. The appendix describes the preprocessing steps, which were data access; mosaicking; raster extraction and cleaning; reprojection, pixel resampling, and cropping; crop mask application; and dataframe construction. At the onset of a crop growing season, input variables are not promptly available on the MODIS data portal due to a delay between data gathering, processing, and online publishing processes. Since our model depends on those variables, we used their historical values to predict potential future trends, most specifically during the crop growing season. For this, a random forest (RF) model was used. RF is a supervised learning model that commonly is used for regression problems. It is also known as the bootstrap aggregator due to its two-step procedure to learn patterns within the dataset: (1) feature and raw sampling with replacement and (2) aggregation with majority-vote rule.

For the RF model used to forecast input variable values during the growing season, the Python open-source sklearn

FIGURE 9.1—AN OVERVIEW OF THE AFCP MODEL DATA PROCESSING AND COMPUTATIONS



Source: Authors.

RF regressor was used. One decision tree (DT) layer was created with 2,000 blocks.

The Africa Food Crop Production Model Computational Aspects

The supervised learning ANNs method was used to build the AfCP model. The inputs were the first four biogeophysical parameters listed in Table 9.2, and the corresponding outputs were the production values (Figure 9.1). A preprocessing data stage dealt with building the proper format and splitting the data into training, validation, and testing sets. The learning process was carried out by building the relationship between inputs and response variables with the training dataset that corresponded to 80 percent of the initial dataset. The validation data (10 percent of the dataset) were used to fine-tune the model. Finally, the testing data (the remaining 10 percent of the dataset) allowed assessment of the model accuracy. The model accuracy was assessed by comparing the model predictions on the test set with the corresponding actual values. This choice was motivated by the unavailability of ground-truth data at the same pixel level. An overall arithmetic average of out-of-sample root-mean-squared error (RMSE) of 0.044 was obtained for all African countries (see Table 9A.6 in the appendix for a detailed overview of the model RMSE across countries and crops). In addition, the aggregated prediction data were compared with public databases such as FAOSTAT (FAO 2021) or food crop production as the sanity check for the AfCP model predictions.

Forecasts were made before the harvesting period (in 2020) for each of the targeted crops (Table 9.3). The FAO crop calendar (FAO n.d.) was used to identify sowing, growing, and harvesting periods. For the eastern African region, cassava and sugarcane were not considered in this study due to data availability issues. For each country [4] and crop, the food crop production modeling work starts at the onset of crop greenness, which is considered the beginning of the growing season. Most of the biogeophysical parameters were not available at that time; therefore, their historical values were used in an RF regressor to estimate their future values in the growing season. The future values were then used as inputs in the food crop production model (see Section 3.3.1).

Food Crop Production Forecasts During the Pandemic

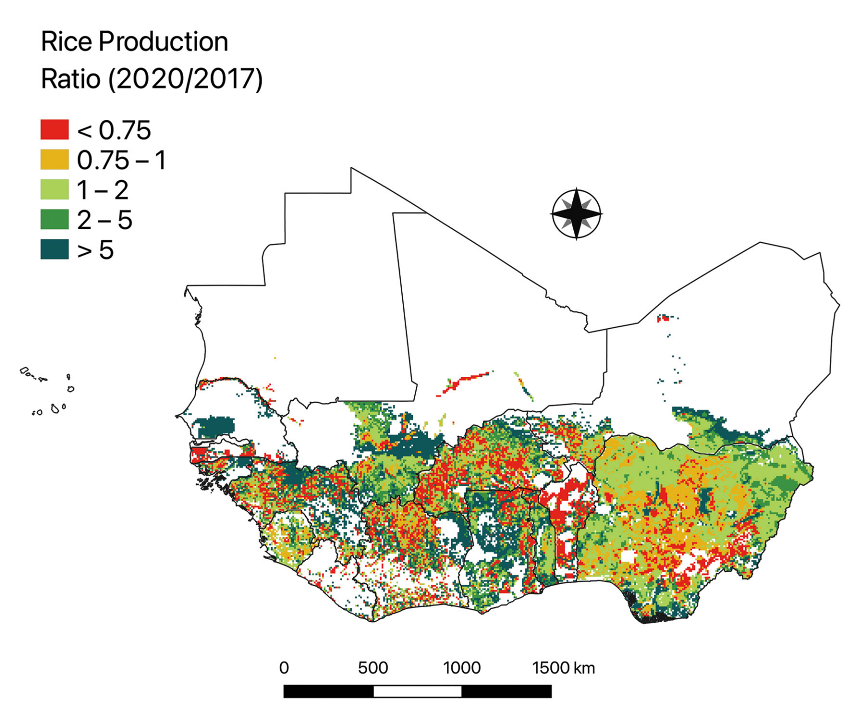
The food crop production model was applied to all African countries and selected crops. Figures 9.2, 9.3, and 9.4 show the 2020 maps' predicted production as a share of the 2017 production for rice, maize, and cassava, respectively, for the western African countries. Figures 9.5 and 9.6 show the maize production ratio for eastern and southern Africa, respectively. Figures 9.7 and 9.8 show wheat and cassava production ratios for northern and central African countries. Figures 9A.4 and 9A.5 in the appendix show the AfCP model's outputs for western (rice, maize, cassava), eastern (maize), northern (wheat), southern (maize), and central (cassava) African regions, respectively.

At the regional level, the production quantities for most of the selected crops in each African region are expected to decline in 2020 as compared to 2017. Only the production quantities for cassava in the western (Figure 9.4) and central (Figure 9.8) African regions are expected to increase, compared with 2017, by 4.2 percent and 28.4 percent, respectively. The sharpest decline in production quantities for the three selected crops in the western African region is expected for rice, with a decrease close to 12 percent, while maize production is expected to decline by close to 5 percent. The decline in maize production is expected to be around 1.5 percent and 18.6 percent for the eastern and southern African regions. Wheat production shows a decline of close to 10 percent in the northern African region in 2020 compared with 2017.

According to the FAO, the six central African countries' aggregated cassava production was around 47 million metric tons in 2017 (FAO 2021). The most significant contributor was the Democratic Republic of the Congo (66.6 percent), followed by Angola (17.9 percent) and Cameroon (10.2 percent). Our model suggests a total cassava production of close to 60 million metric tons for the same countries in 2020, which corresponds to an increase of 28 percent compared to 2017. However, in 2020, the distribution of total production across individual countries is expected to remain the same for Angola, while

4 Food crop production forecasts were performed for each African country. Then national maps were merged together to obtain regional maps. Such a strategy is justified by the need to avoid the so-called "ecological fallacy," which in this case means making predictions for a country based on an aggregated dataset at a continental level. The consequence of that would be country input data impacting the output data for another country.

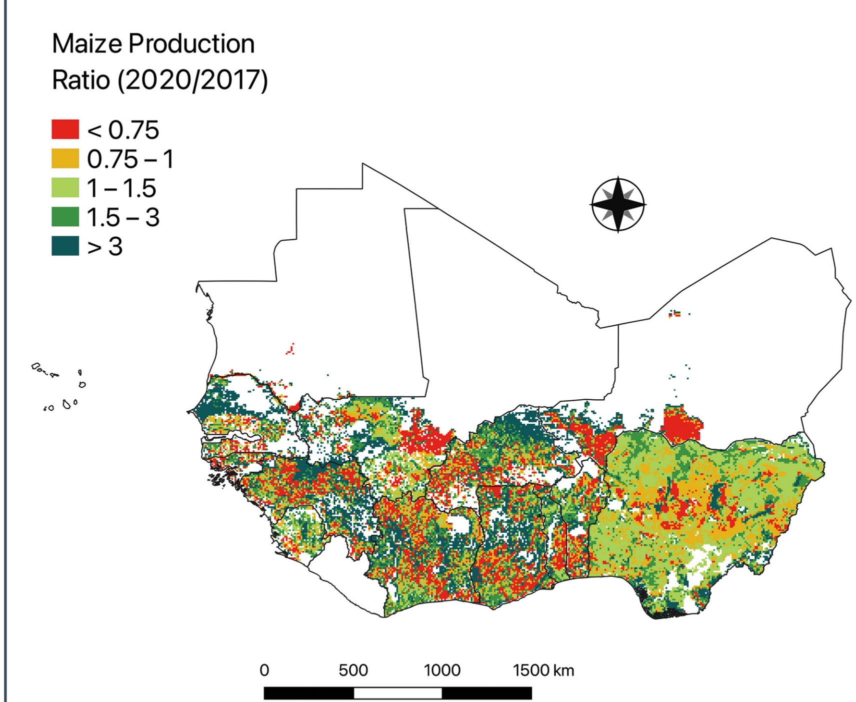
FIGURE 9.2—THE 2020 PREDICTED RICE PRODUCTION AS A SHARE OF THE 2017 PRODUCTION FOR WESTERN AFRICA



Source: Authors.

Note: If the ratio is above unity, the 2020 predicted production is expected to be larger than the 2017 production. A ratio smaller than unity means an expected decrease in production in 2020 compared to 2017.

FIGURE 9.3—THE 2020 PREDICTED MAIZE PRODUCTION AS A SHARE OF THE 2017 PRODUCTION FOR WESTERN AFRICA



Source: Authors.

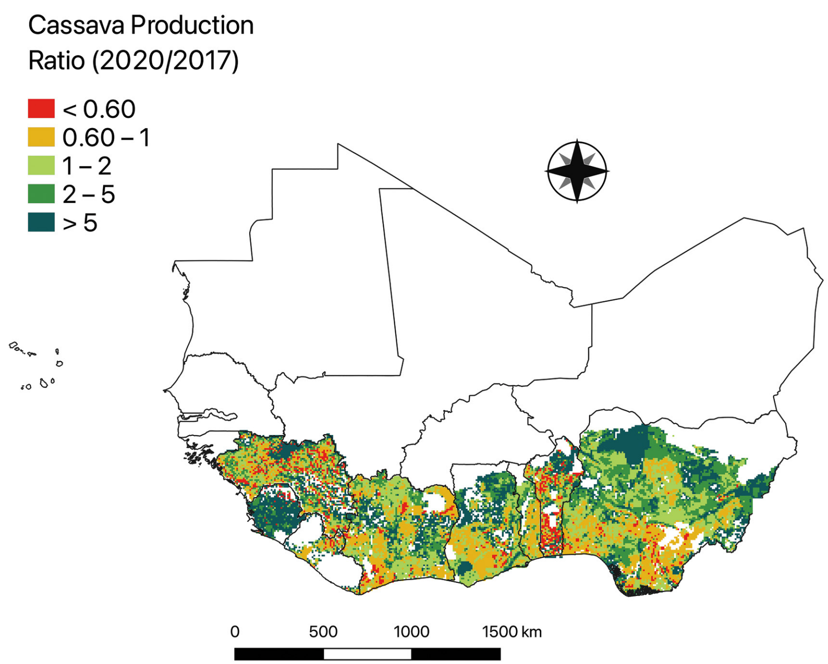
Note: If the ratio is above unity, the 2020 predicted production is expected to be larger than the 2017 production. A ratio smaller than unity means an expected decrease in production in 2020 compared to 2017.

Cameroon's shares are expected to decrease by roughly half (5.6 percent). The Democratic Republic of the Congo has a share increase of nearly 10 percent.

In northern Africa, the ratios between predicted wheat production levels for the 2020 season and actual levels in 2017 show a slight decrease in 2020. On average, the map (Figure 9.7) suggests better wheat production for the 2017 season compared with 2020 for Libya, Tunisia, Morocco, and Algeria. Compared with 2017 production levels, projected wheat production in 2020 in Sudan and Egypt shows an increase of 3.5 percent and 2.8 percent, respectively, while other countries show a decline.

The aggregated maize production predicted from our model for eastern countries is around 28 million metric tons in 2020. Each country's contribution to the global production is as follows: Ethiopia (27.7 percent), Tanzania (20.2 percent), Kenya (12.2 percent), Zambia (10.9 percent), Uganda (10.7 percent), Malawi (9.1 percent), Mozambique (5 percent), Zimbabwe (2.5 percent), Rwanda (1.1 percent), and Madagascar (0.8 percent). In 2017, the production was estimated at around 28.5 million metric tons for the same countries (IFPRI 2020). There is a slight decrease of 1.6 percent in 2020 production estimates compared with 2017. However, some countries, such as Ethiopia, Zimbabwe, Uganda, and Zambia, show a slight increase in their production of

FIGURE 9.4—THE 2020 PREDICTED CASSAVA PRODUCTION AS A SHARE OF THE 2017 PRODUCTION FOR WESTERN AFRICA



Source: Authors.

Note: If the ratio is above unity, the 2020 predicted production is expected to be larger than the 2017 production. A ratio smaller than unity means an expected decrease in production in 2020 compared to 2017.

0.8 percent, 13.2 percent, 4.6 percent, and 3.8 percent, respectively, compared with 2017.

Figure 9.6 shows ratios between predicted maize production levels for the 2020 season and actual levels in 2017 in southern Africa. The map suggests disparities in maize production for the 2020 season compared with 2017. Maize production is expected to decline by 30 percent in South Africa, 20 percent in Lesotho, 4.9 percent in Eswatini, and 0.7 percent in Namibia from 2017 to 2020.

While the biogeophysical parameters might not allow the establishment of a causal effect between the agricultural production variabilities and COVID-19, the gap between FAO agricultural production data and the AfCP model

TABLE 9.4—TOTAL PRODUCTION IN 2017 AND 2020, AND THE RATE OF CHANGE FOR EACH CROP AND AFRICAN REGION

Region	Crop	2017 Production (MT)	2020 Predicted Production (MT)	Rate of change (%)
Western Africa	Rice	17,803,495.8	15,640,125.8	−12.15
	Maize	21,666,866.9	20,599,545.5	−4.92
	Cassava	90,151,658.8	93,948,433.2	+4.21
Eastern Africa	Maize	28,539,928.7	28,095,011.8	−1.55
Northern Africa	Wheat	18,392,407.2	16,610,688.1	−9.68
Southern Africa	Maize	420,814.5	342,688.3	−18.56
Central Africa	Cassava	47,209,110.0	60,598,537.0	+28.36

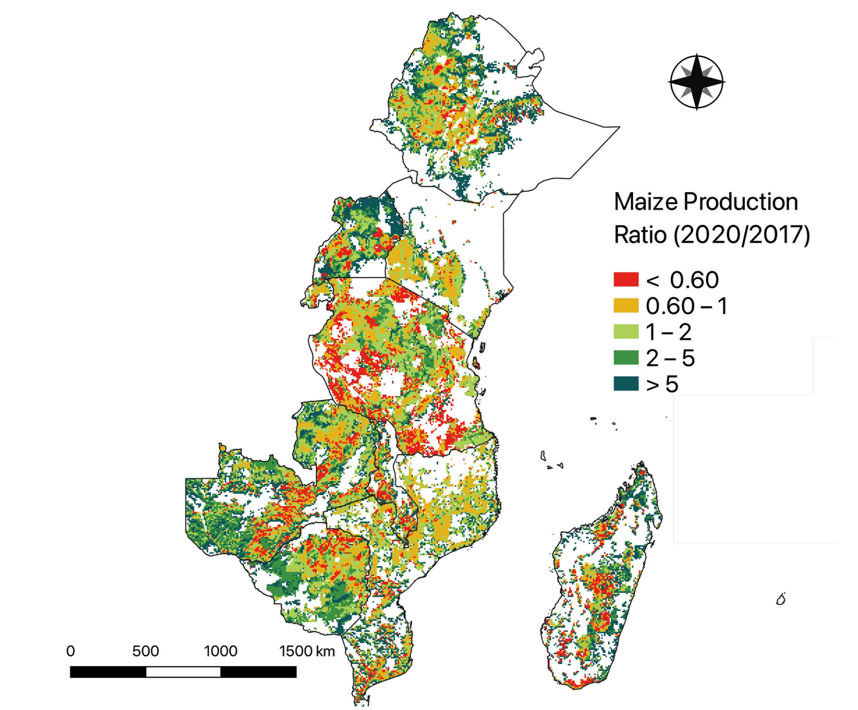
Source: For 2017 production, FAOSTAT; for 2020 production, authors.

Note: MT = metric tons.

predictions can provide information about the potential impacts of government restriction measures.

Indeed, COVID-19 was declared a pandemic on March 11, 2020, and lockdown measures followed shortly afterward. The onset of the pandemic in Africa coincided with the seed marketing period for maize in West Africa and with the end of cassava seed marketing for the same region (de Boef et al. 2021). Seed scarcity is a consequence of mobility restriction measures, and it significantly affects food crop production and pushes farmers to rely more on the informal seed market, leading to less potential for high yields. From the same study, 79 percent of panelists from Myanmar, Nigeria, Ethiopia, and Uganda reported facing significant difficulty in obtaining sufficient volumes of quality early generation seeds of desired varieties. Moreover, according to a forecast from the national seed committees of member states of the Economic Community of West African States and the Permanent Interstate Committee for Drought Control in the Sahel, there will be a shortfall of certified seeds for maize and millet in the 2020 cropping season (CORAF 2020). Fewer than 10,000 metric tons of certified sorghum and millet seeds were produced in 2020 compared with a demand of about 100,000 metric tons, representing only about 10 percent availability. As for maize, about 70,000 metric tons are available, though the need is close to 200,000 metric tons (CORAF 2020). The lack

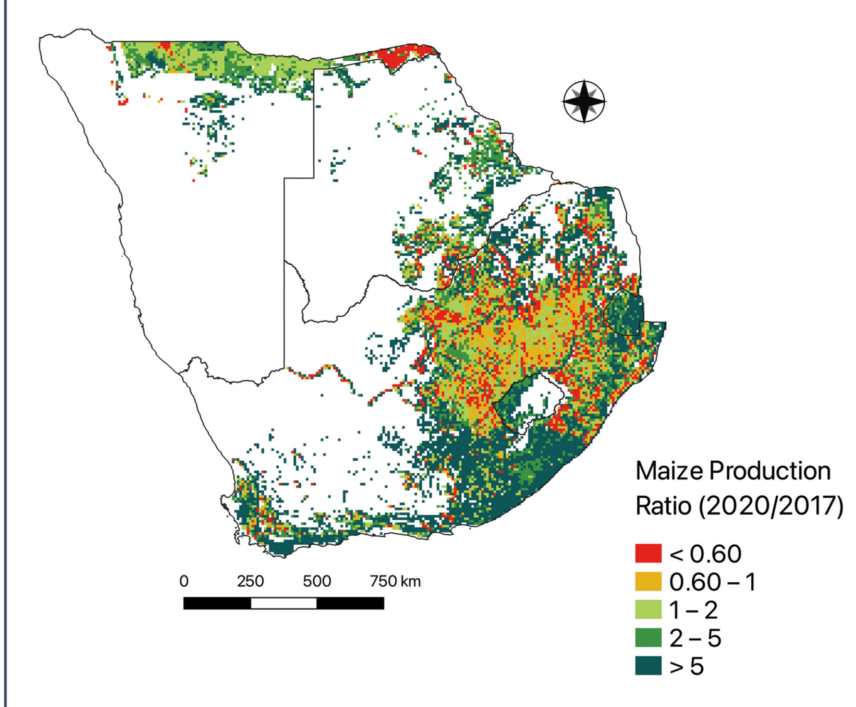
FIGURE 9.5—THE 2020 PREDICTED MAIZE PRODUCTION AS A SHARE OF THE 2017 PRODUCTION FOR EASTERN AFRICA



Source: Authors.

Note: If the ratio is above unity, the 2020 predicted production is expected to be larger than the 2017 production. A ratio smaller than unity means an expected decrease in production in 2020 compared to 2017.

FIGURE 9.6—THE 2020 PREDICTED MAIZE PRODUCTION AS A SHARE OF THE 2017 PRODUCTION FOR SOUTHERN AFRICA



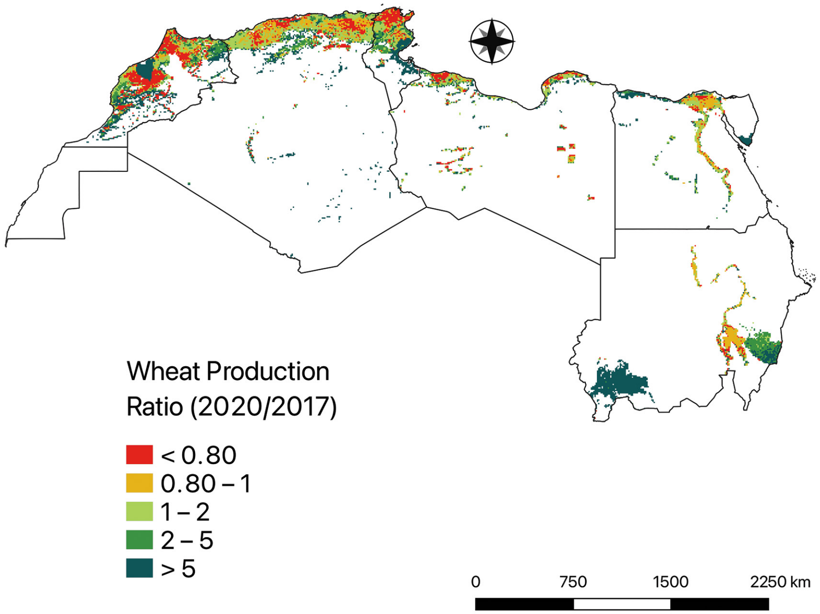
Source: Authors.

Note: If the ratio is above unity, the 2020 predicted production is expected to be larger than the 2017 production. A ratio smaller than unity means an expected decrease in production in 2020 compared to 2017.

of timely reception and use of seeds due to mobility restriction measures can explain the food crop production decrease at the regional level. However, each country has its own COVID-19 policy for agricultural inputs and a different strategy to secure seeds for its national farmers. This could also explain why some countries in the same region show an increase in the production of a specific crop, while others show a decrease. However, only a comparative analysis between countries' policies could help confirm this assertion. The seed scarcity also coincides with an increase in transaction costs that hinder farmers' access to affordable and certified seeds.

The mobility restriction measures implemented to reduce the spread of COVID-19 have also affected labor mobility, especially for a labor-intensive sector such as agriculture. The sector employs 70 percent (World Bank 2013) of the total workforce in the region south of the Sahara. The planting period corresponds to the peak of labor demand; therefore, any disruption in labor supply will potentially have a negative impact on food crop production. Moreover, most countries implemented border closures during the planting period of maize and rice, with the consequence of potentially delaying the harvesting period. There is a high probability that mobility restriction measures and labor scarcity will significantly impact the production of major staple crops in the region.

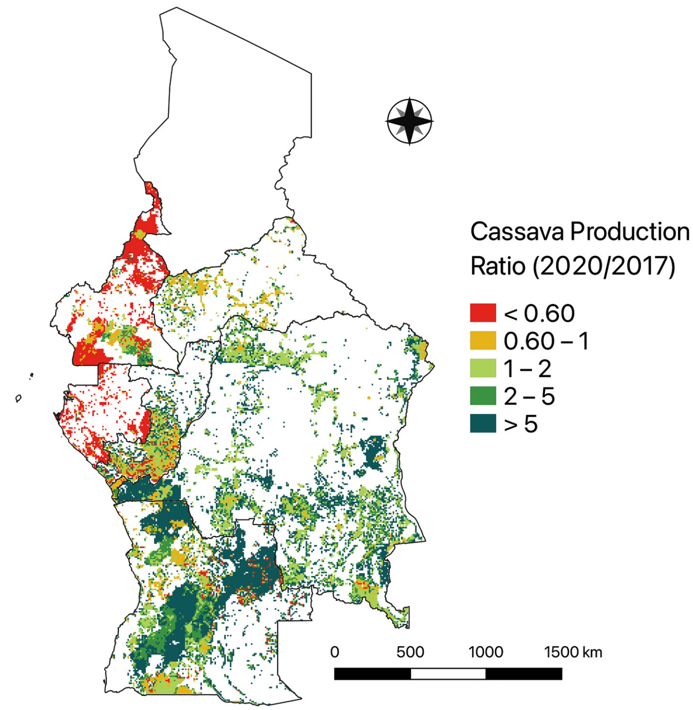
FIGURE 9.7—THE 2020 PREDICTED WHEAT PRODUCTION AS A SHARE OF THE 2017 PRODUCTION FOR NORTHERN AFRICA



Source: Authors.

Note: If the ratio is above unity, the 2020 predicted production is expected to be larger than the 2017 production. A ratio smaller than unity means an expected decrease in production in 2020 compared to 2017.

FIGURE 9.8—THE 2020 PREDICTED CASSAVA PRODUCTION AS A SHARE OF THE 2017 PRODUCTION FOR CENTRAL AFRICA



Source: Authors.

Note: If the ratio is above unity, the 2020 predicted production is expected to be larger than the 2017 production. A ratio smaller than unity means an expected decrease in production in 2020 compared to 2017.

As Figures 9.2–8 show, while many areas are expected to experience declines in 2020 production compared with 2017 levels, production is predicted to be higher in other areas. In addition to the impacts of COVID-19, many other factors affect crop production in complex ways: climate and weather variability in particular drive much of the variability in crop production.

In addition to the potential to monitor crops' growing conditions through biogeophysical parameters, the combination of RSPs and machine learning provides several other benefits. The RSPs allow us to bring disaggregation to the community level while the machine learning techniques help us predict food crop production before the harvesting period. These two outcomes are valuable assets to strengthening food production systems in Africa through improved agricultural statistics and analytics. However, the path from RSPs and machine learning to policymaking in the agriculture sector requires several steps.

The AfCP Model Limitations and Future Direction of Research

Nonbiogeophysical Input Parameters

The rationale for building the AfCP model was to provide data and forecasts about agricultural production to help navigate the uncertainties of COVID-19 in the African agricultural sector. We also encountered the need to further disaggregate the data due to the localized impacts of such a crisis (or any crisis), which sometimes are either not perceived or are weakly perceived at the national level. In such a process, the choice of biogeophysical data for the AfCP was made for three main reasons: (1) they provide disaggregated data by nature, (2) they have sufficiently long time-series data for machine learning-type predictive modeling, and (3) they have direct effects on food crop production. However, we acknowledge that other nonbiogeophysical parameters can also have a significant impact on food production, such as population density (as a proxy for labor) and food security status. Such aspects will be included in a future version of the AfCP model.

Spatial Resolution

Another limitation of the AfCP is its spatial resolution, which is 10 kilometers. Such a low resolution is explained by the crop masks that were derived from SPAM production rasters. Because they were the lowest resolution maps among the input and response variables, the choice was made to aggregate the highest resolution rasters' pixels to 10 kilometers rather than resampling the crop masks at a higher resolution, which would bring significant homogenization to the dataset and lower the learning capacity of the AfCP. While the 10-kilometer pixel values are representative of the area covered, they are much larger than typical African farm sizes of fewer than 2 hectares.

An improvement to the AfCP would be the use of higher spatial-resolution rasters, which would allow the further disaggregation of the predictions and increase the overall accuracy of the model. Such an improvement is ongoing, currently with a crop mask derived by the recently released cropland map at 10 meters spatial resolution from the Environmental Systems Research Institute, based on data from Sentinel-2. However, the current degree of spatial resolution already allows us to predict production at a more disaggregated level than the usual administrative divisions in Africa and, therefore, to generate evidence to inform specific policies at the community level.

Ground Truthing

Another limitation of the AfCP is that it was not calibrated with ground-truth data. The data that would allow for that are not available at the same spatial resolution. Therefore, the accuracy assessment was performed on the test set. There is a need to improve the model accuracy with countries' data even though a comparison between predicted and actual values from public datasets, such as the FAOSTAT portal, yielded good results at subnational and national levels. AKADEMIYA2063 is working toward building the partnerships that would allow ground truthing.

Conclusion

As suggested by the impacts of COVID-19, a robust African agricultural statistics system is much needed to create informed and targeted responses and policies. Building a culture of gathering accurate and timely data about features related to food crop production would not only facilitate the production of better policies and monitoring and evaluation mechanisms, it would also be critical to increasing countries' levels of preparedness for any potential future crisis in the sector. Data gathering could help improve preparedness by identifying the crisis early enough to mitigate its impacts or by allowing decision-makers and policymakers to better manage it. This chapter explains how emerging technologies such as RSPs and machine learning can be harnessed to provide valuable information for decision-making processes in the agricultural sector. The AfCP model has been in development from the onset of the COVID-19 crisis. Although the pandemic has been the pretext for the predictive modeling work, this sort of model architecture could be used or adapted outside of the COVID-19 context. However, it is worth noting that any adaptation will require carefully choosing the explanatory variables and ensuring their availability at the pixel level. The path from a raw satellite image to an informative map is not straightforward; several areas of expertise need to operate at different levels.

Capacity building for emerging technologies such as remote sensing and machine learning should be institutionalized. African governments must create special units in which emerging technologies can be harnessed to inform policies. Moreover, incentivizing initiatives in the private sector to do the same would also benefit African countries. However, attracting students into the fields of emerging technologies requires both sectors to create solid public-private partnerships and support for entrepreneurship in science, technology, engineering, and mathematics to create jobs and ensure the availability of a critical mass of experts.

Across African countries, data related to food crop production in the agricultural sector are collected at the season onset and after the harvesting period. The data are mainly collected at the household level, including household information, crop type, production quantities, land size, availability and use of agricultural inputs, post-harvest loss, and trade information. This methodology has proven sufficient for an extended period. However, given technological improvements and their use in more efficient data-gathering processes, there is a need to take full advantage of the current data-gathering technologies, such as unmanned aerial vehicles for spectral signature data gathering and monitoring purposes or tablets with predefined surveys combined with cloud infrastructure for storage and computations.

Metadata are as essential as primary data in obtaining the benefits of recent analytics tools and predictive modeling through machine learning. Metadata help to contextualize the primary datasets and add more explanatory variables into the predictive model for more robustness. The use of cloud technology, telecommunications, and tablets with embedded optimized forms could facilitate gathering such third-party information. The cloud would help to store the data and perform further analysis; the Internet connection could help gather GPS coordinates and inform about the locations where the data were gathered (not only at the household level but at the farm itself, allowing the analysis of biogeophysical parameters from RSPs). For this to occur, at least three enabling technologies are required: (1) the improvement of Internet connections in rural areas where most farms are located, (2) the inclusion of metadata information gathering into agricultural surveys, and (3) the renewal of data-gathering tools to migrate data from papers and laptops to tablets that are more suitable for such a task. Such an approach of using emerging and well-established technologies to support better-quality data gathering in the agricultural sector will progressively require fewer resources, because the use of remote sensing will reduce the need to update some data from the ground.

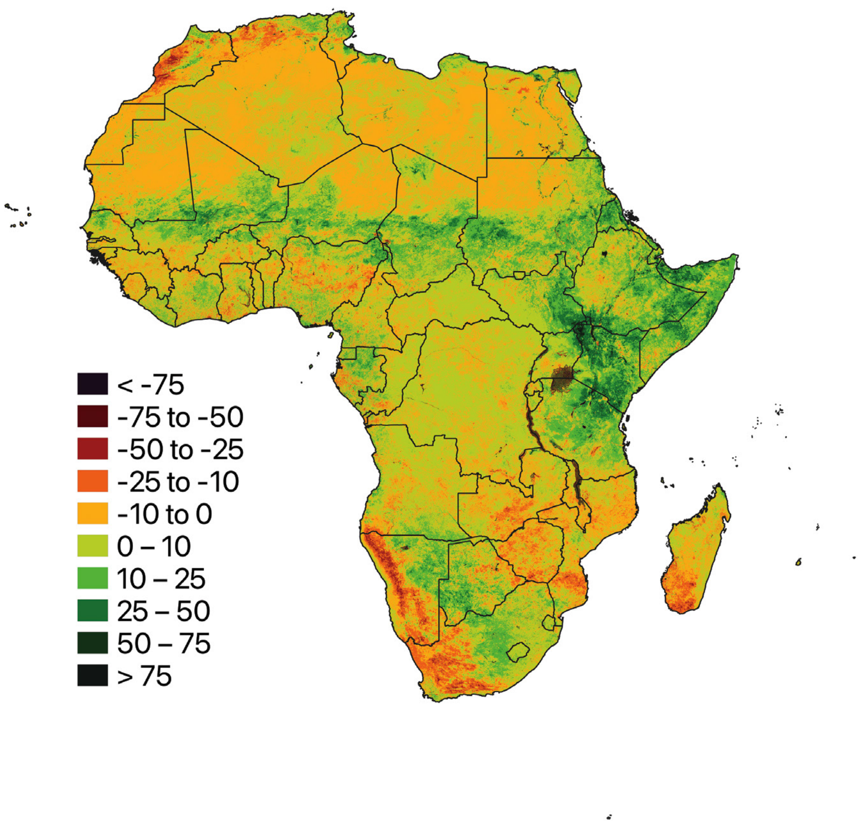
Information asymmetry between researchers and policymakers is a long-standing problem in Africa, especially in the agricultural sector. Moreover, the fast pace of turnover in offices makes the consolidation of technical knowledge within an institution difficult. For instance, an individual at a national statistics bureau could be trained to work with remote sensing products and machine learning techniques within a year. The following year, that individual could have moved to another ministry, another entity of the same ministry, or another institution. From a general point of view, the training is not lost. However, the corresponding technical capacity moves from one entity to another with the risk that is not used where most needed.

The complex African cropping system makes it difficult to collect accurate and timely data in a sustainable way. Data scarcity does not allow the type of detailed analysis that decision-making requires in a time of uncertainty. Even when the data quality and disaggregation requirements are met, however, the way the knowledge is produced seems to be inaccessible to policymakers, especially when emerging technologies are used and are far from reach. One way of closing this gap is to use data visualization expertise to transform data and knowledge from a raw stage to an informational stage. Such expertise is not yet well developed across African countries and needs to be built.

The results of this chapter not only support the use of emerging technologies such as RSPs and machine learning techniques to improve agricultural statistics, but also show how they could be leveraged to increase African countries' preparedness to shocks after COVID-19. The pandemic has shown how much timely and accurate data are needed for early action and intervention in the agricultural sector and beyond. Recent technologies must be considered in every part of the data environment—from collection to analysis.

Appendix

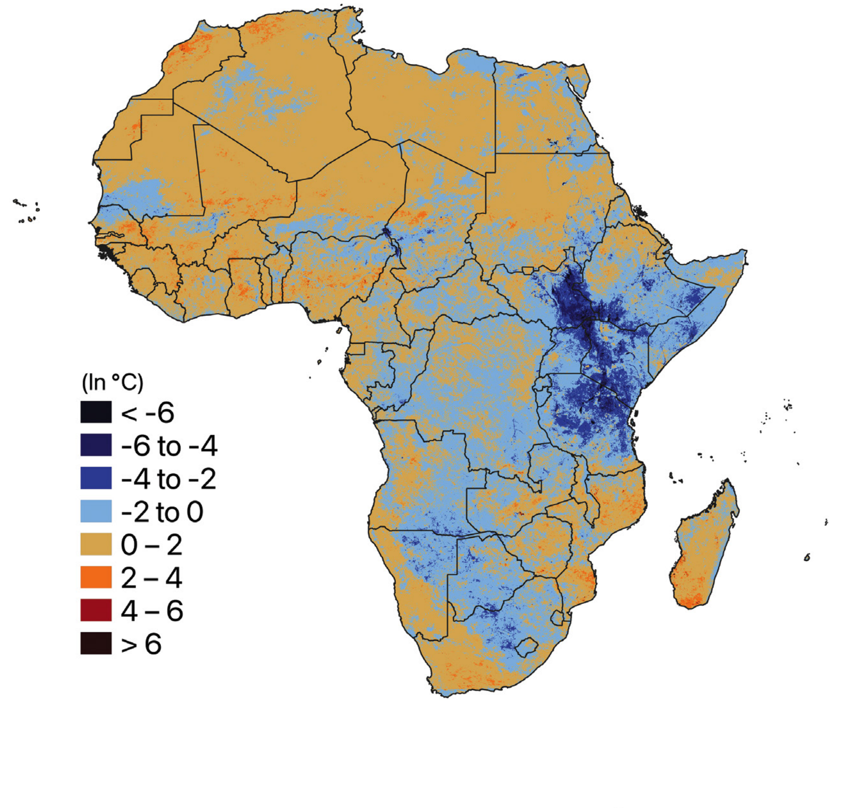
FIGURE 9A.1—NDVI ANOMALY IN AFRICA FOR THE YEAR 2020



Source: Data processing and map from authors.

Note: The values provided are relative to the twenty years (2000–2019) average for each pixel.

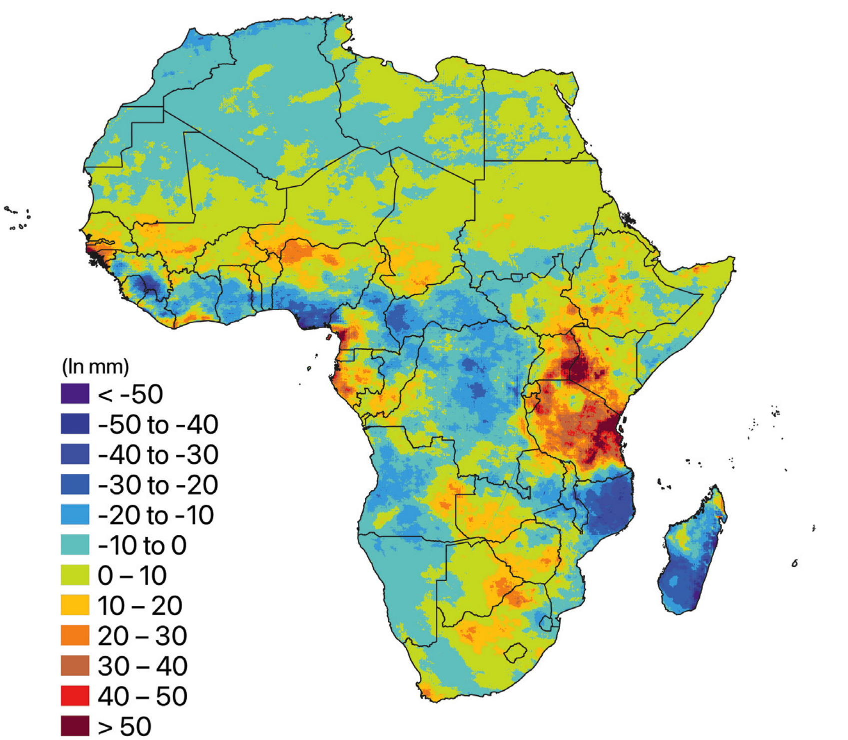
FIGURE 9A.2—DAYTIME LST ANOMALY IN AFRICA FOR THE YEAR 2020



Source: Data processing and map from authors.

Note: The values provided are the absolute difference between the 2020 pixel values and their historical average (2000–2019).

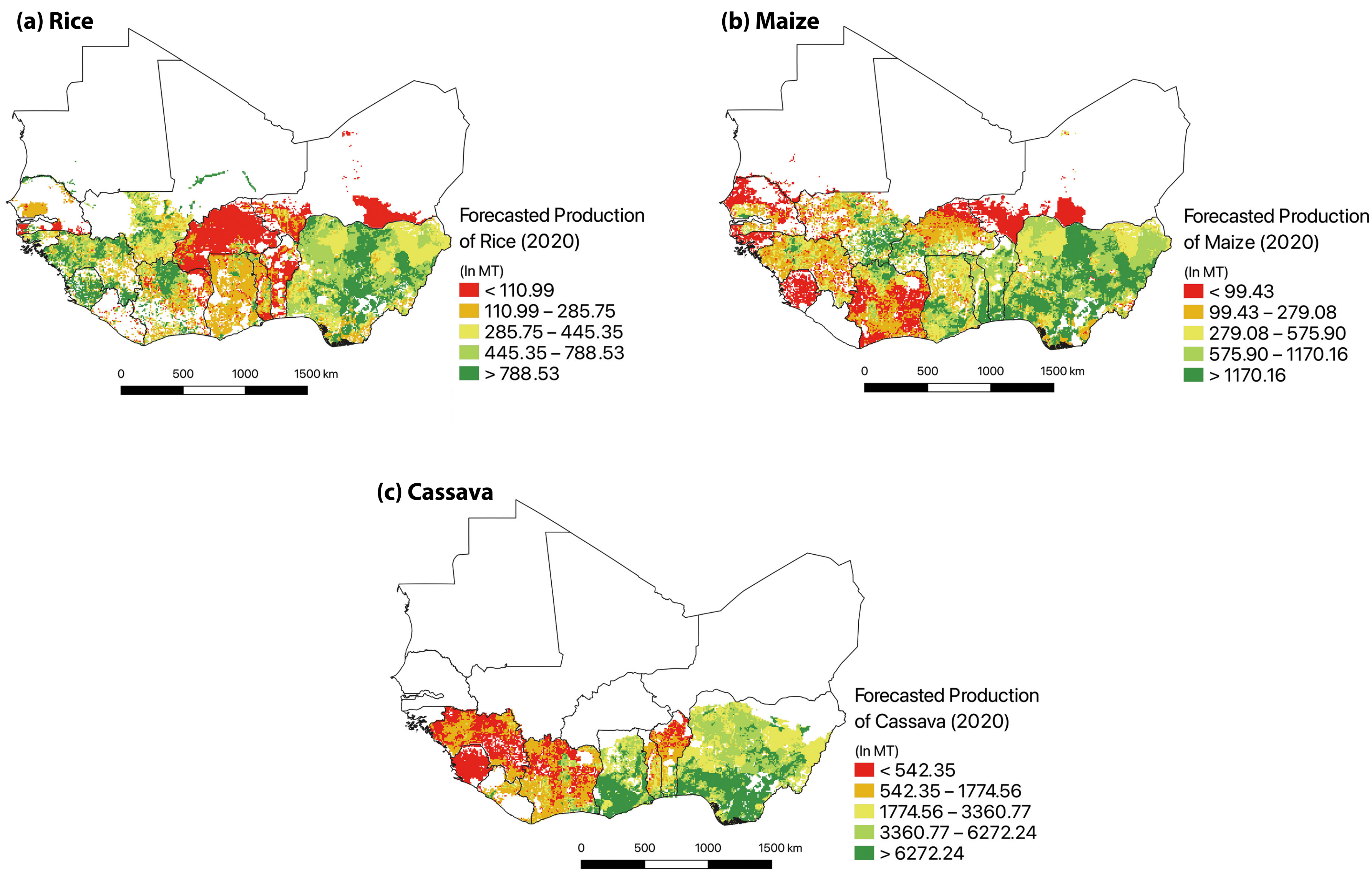
FIGURE 9A.3—RAINFALL ANOMALY IN AFRICA FOR THE YEAR 2020



Source: Data processing and map from authors.

Note: The values provided are the absolute difference between the 2020 pixel values and their historical average (2000–2019).

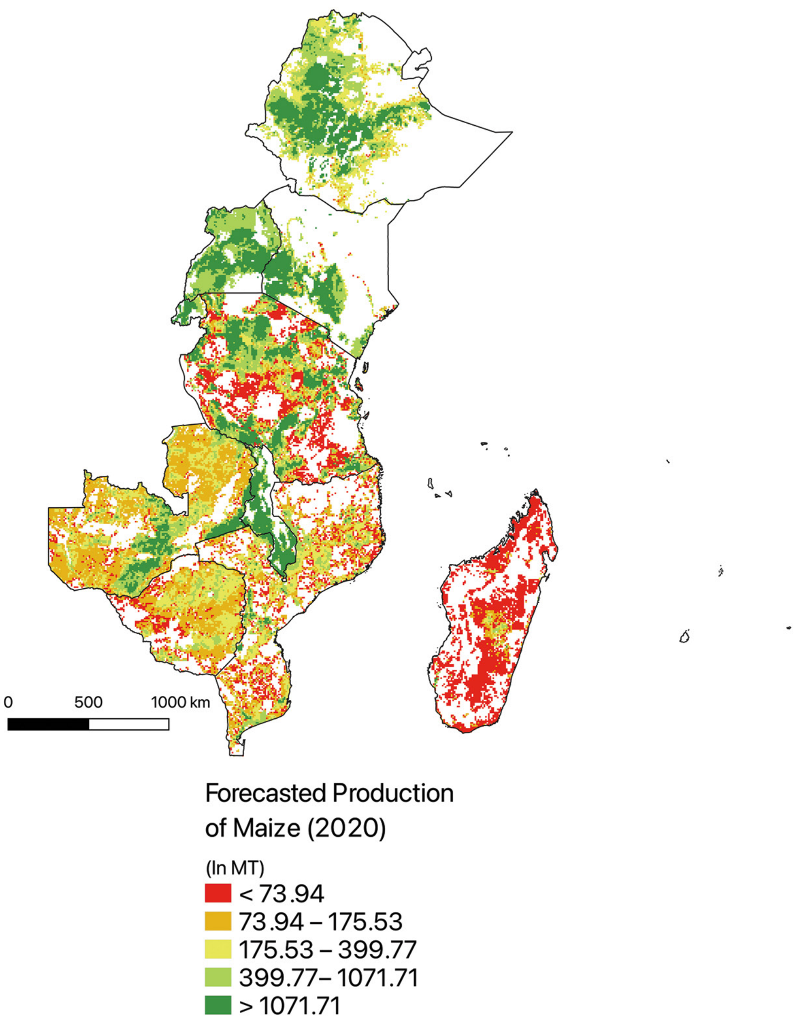
FIGURE 9A.4—THE 2020 PREDICTED (A) RICE, (B) MAIZE, AND (C) CASSAVA PRODUCTION IN WESTERN AFRICAN COUNTRIES



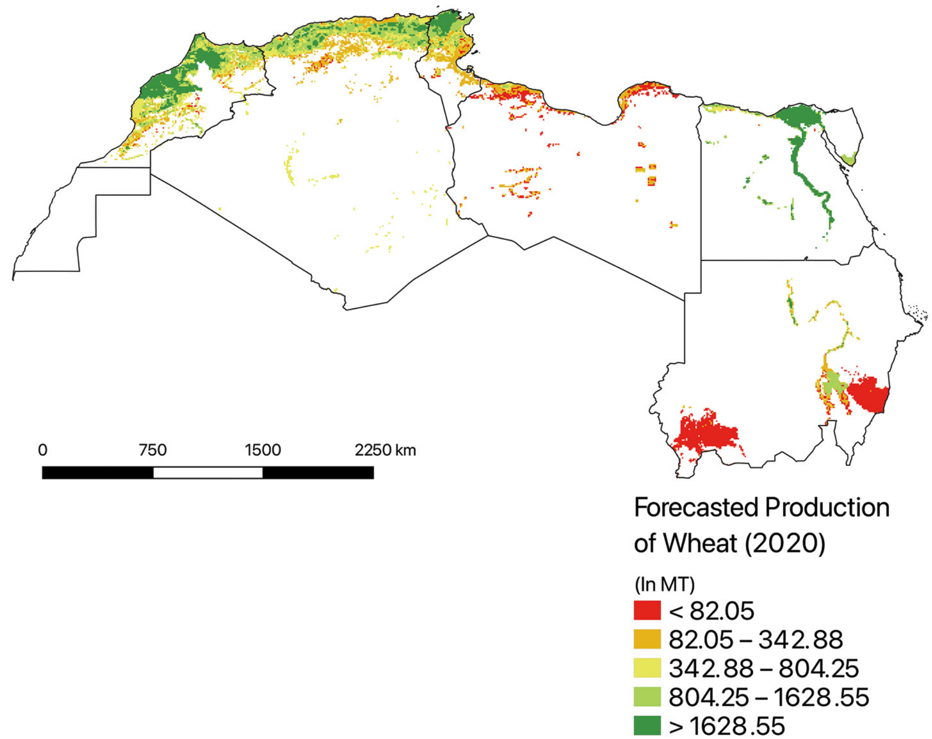
Source: Data, methodology, and maps from authors.

FIGURE 9A.5— THE 2020 PREDICTED (A) MAIZE PRODUCTION IN EASTERN AFRICAN COUNTRIES, (B) WHEAT PRODUCTION IN NORTHERN AFRICAN COUNTRIES, (C) MAIZE PRODUCTION IN SOUTHERN AFRICAN COUNTRIES, AND (D) CASSAVA PRODUCTION IN CENTRAL AFRICAN COUNTRIES *(continued on next page)*

(a) Maize—eastern African countries



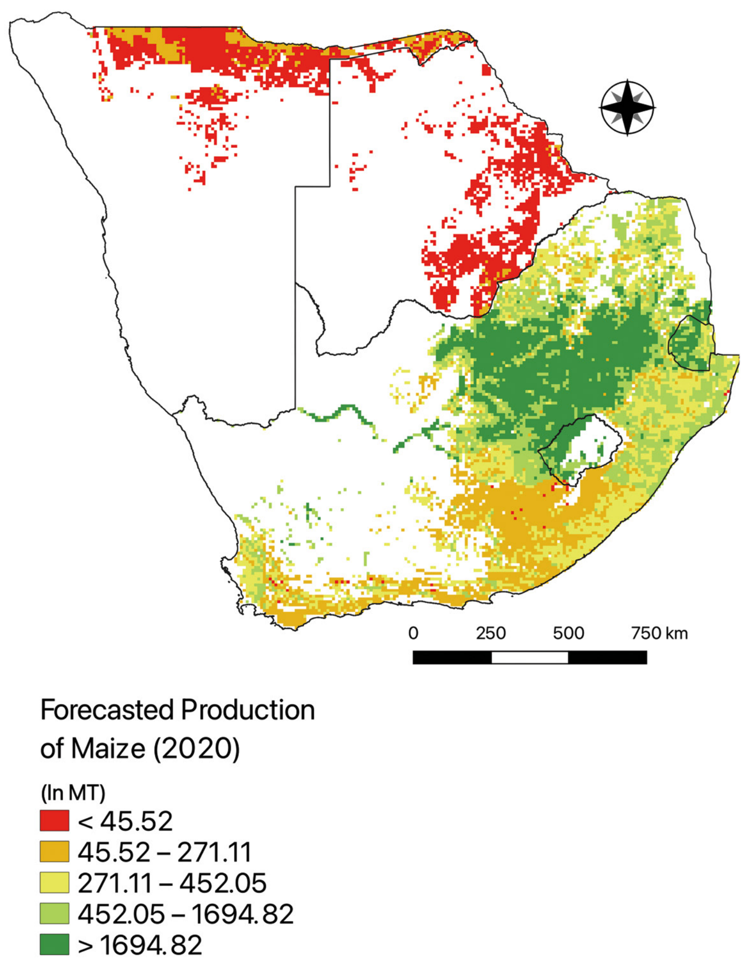
(b) Wheat—northern African countries



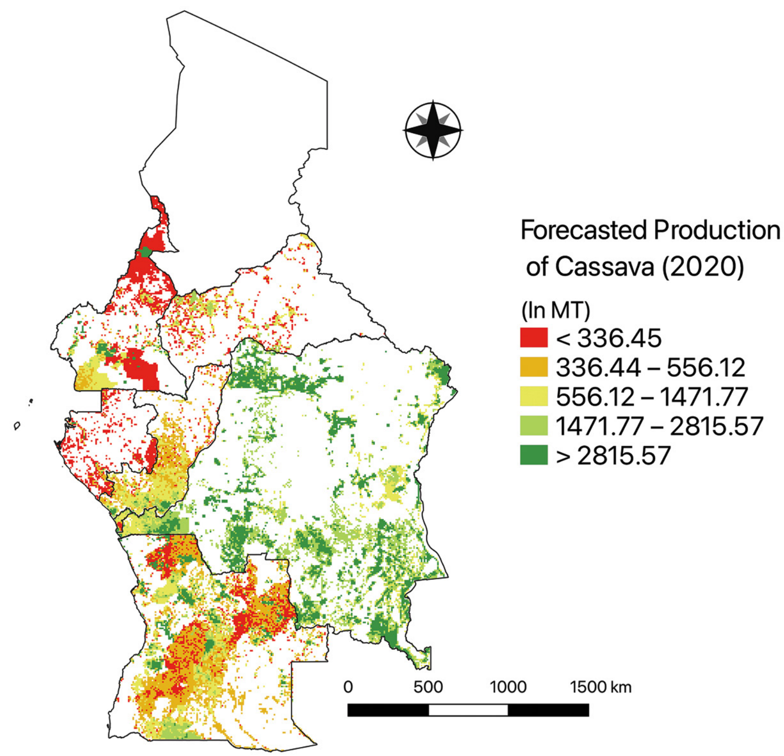
Source: Data, methodology, and maps from authors.

FIGURE 9A.5— THE 2020 PREDICTED (A) MAIZE PRODUCTION IN EASTERN AFRICAN COUNTRIES, (B) WHEAT PRODUCTION IN NORTHERN AFRICAN COUNTRIES, (C) MAIZE PRODUCTION IN SOUTHERN AFRICAN COUNTRIES, AND (D) CASSAVA PRODUCTION IN CENTRAL AFRICAN COUNTRIES *(continued from previous page)*

(c) Maize—southern African countries



(d) Cassava—central African countries



Source: Data, methodology, and maps from authors.

Appendix *continued*

TABLE 9A.1—MOST SIGNIFICANT COMMODITIES BY COUNTRY IN WESTERN AFRICA REGION BASED ON OUR SELECTION CRITERIA

Country	Commodity 1	Commodity 2	Commodity 3	Commodity 4	Commodity 5
Benin	Cassava and products	Yams	Maize and products	Beverages, fermented	Palm kernels
Burkina Faso	Beverages, fermented	Sorghum and products	Maize and products	Millet and products	Pulses, other, and products
Cabo Verde	Sugarcane	Pelagic fish	Tomatoes and products	Vegetables, other	Milk, excluding butter
Côte d'Ivoire	Yams	Cassava and products	Rice and products	Palm kernels	Sugarcane
Gambia	Groundnuts (shelled equivalent)	Millet and products	Milk, excluding butter	Rice and products	Beverages, fermented
Ghana	Cassava and products	Yams	Plantains	Palm kernels	Maize and products
Guinea	Rice and products	Cassava and products	Palm kernels	Maize and products	Groundnuts (shelled equivalent)
Guinea-Bissau	Rice and products	Nuts and products	Roots, other	Palm kernels	Plantains
Liberia	Cassava and products	Rice and products	Sugarcane	Palm kernels	Bananas
Mali	Maize and products	Rice and products	Millet and products	Vegetables, other	Milk, excluding butter
Mauritania	Pelagic fish	Milk, excluding butter	Rice and products	Demersal fish	Sorghum and products
Niger	Millet and products	Pulses, other, and products	Sorghum and products	Vegetables, other	Milk, excluding butter
Nigeria	Cassava and products	Yams	Vegetables, other	Maize and products	Palm kernels
Senegal	Sugarcane	Groundnuts (shelled equivalent)	Rice and products	Millet and products	Vegetables, other
Sierra Leone	Cassava and products	Rice and products	Vegetables, other	Palm kernels	Milk, excluding butter
Togo	Cassava and products	Maize and products	Yams	Sorghum and products	Beans

TABLE 9A.2—MOST SIGNIFICANT COMMODITIES BY COUNTRY IN NORTHERN AFRICA REGION BASED ON OUR SELECTION CRITERIA

Country	Commodity 1	Commodity 2	Commodity 3	Commodity 4	Commodity 5
Algeria	Vegetables, other	Potatoes and products	Milk, excluding butter	Wheat and products	Onions
Egypt	Sugarcane	Sugar beet	Wheat and products	Vegetables, other	Maize and products
Morocco	Wheat and products	Sugar beet	Vegetables, other	Milk, excluding butter	Barley and products
Sudan	Sugarcane	Sorghum and products	Milk, excluding butter	Groundnuts (shelled equivalent)	Onions
Tunisia	Vegetables, other	Milk, excluding butter	Tomatoes and products	Wheat and products	Olives (including preserved)

TABLE 9A.3—MOST SIGNIFICANT COMMODITIES BY COUNTRY IN SOUTHERN AFRICA REGION BASED ON OUR SELECTION CRITERIA

Country	Commodity 1	Commodity 2	Commodity 3	Commodity 4	Commodity 5
Botswana	Milk, excluding butter	Beer	Roots, other	Vegetables, other	Bovine meat
Eswatini	Sugarcane	Sugar (raw equivalent)	Alcohol, non-food	Maize and products	Roots, other
Lesotho	Milk, excluding butter	Potatoes and products	Maize and products	Beer	Vegetables, other
Namibia	Roots, other	Pelagic fish	Beer	Demersal fish	Milk, excluding butter
South Africa	Sugarcane	Maize and products	Milk, excluding butter	Beer	Potatoes and products

TABLE 9A.4—MOST SIGNIFICANT COMMODITIES BY COUNTRY IN CENTRAL AFRICA REGION BASED ON OUR SELECTION CRITERIA

Country	Commodity 1	Commodity 2	Commodity 3	Commodity 4	Commodity 5
Angola	Cassava and products	Bananas	Maize and products	Sweet potatoes	Beer
Cameroon	Cassava and products	Plantains	Maize and products	Palm kernels	Roots, other
Central African Republic	Cassava and products	Yams	Groundnuts (shelled equivalent)	Roots, other	Sugarcane
Chad	Sorghum and products	Groundnuts (shelled equivalent)	Millet and products	Milk, excluding butter	Cereals, other
Congo	Cassava and products	Sugarcane	Beer	Vegetables, other	Palm kernels
Gabon	Plantains	Cassava and products	Sugarcane	Yams	Beer
Sao Tome and Principe	Plantains	Coconuts, including copra	Palm kernels	Roots, other	Pelagic fish

TABLE 9A.5—MOST SIGNIFICANT COMMODITIES BY COUNTRY IN EASTERN AFRICA REGION BASED ON OUR SELECTION CRITERIA

Country	Commodity 1	Commodity 2	Commodity 3	Commodity 4	Commodity 5
Djibouti	Vegetables, other	Milk, excluding butter	Bovine meat	Mutton and goat meat	Fruits, other
Ethiopia	Maize and products	Roots, other	Cereals, other	Sorghum and products	Wheat and products
Kenya	Sugarcane	Milk, excluding butter	Maize and products	Vegetables, other	Potatoes and products
Madagascar	Rice and products	Sugarcane	Cassava and products	Sweet potatoes	Fruits, other
Malawi	Cassava and products	Sweet potatoes	Maize and products	Sugarcane	Fruits, other
Mauritius	Sugarcane	Sugar (raw equivalent)	Vegetables, other	Poultry meat	Beer
Mozambique	Cassava and products	Sugarcane	Maize and products	Milk, excluding butter	Bananas
Comoros	Pelagic fish	Marine fish, other	Demersal fish	Crustaceans	
Rwanda	Bananas	Sweet potatoes	Cassava and products	Potatoes and products	Plantains
Seychelles	Pelagic fish	Demersal fish	Marine fish, other	Fish, body oil	Aquatic animals, other
Uganda	Sugarcane	Plantains	Cassava and products	Maize and products	Beverages, fermented
United Republic of Tanzania	Maize and products	Cassava and products	Sweet potatoes	Bananas	Sugarcane
Zambia	Sugarcane	Maize and products	Cassava and products	Sugar (raw equivalent)	Milk, excluding butter
Zimbabwe	Sugarcane	Maize and products	Sugar (raw equivalent)	Milk, excluding butter	Cassava and products

TABLE 9A.6—THE AFCP MODEL TRAINING AND ACCURACY ASSESSMENT ACROSS COUNTRIES AND CROPS

Country	Crop	Minimum loss values	Mean loss values	Maximum loss values	RMSE on test set	RMSE on training set
Benin	Maize	0.00169249	0.00433745	0.02844726	0.03255314	0.03179684
Guinea	Maize	0.00242923	0.00383869	0.01885259	0.05860499	0.04483125
Ghana	Maize	0.00249482	0.00436432	0.02395666	0.03866078	0.04119789
Mauritania	Maize	0.00278496	0.00909949	0.04923604	0.03857696	0.03871264
Mali	Maize	0.00198173	0.00337625	0.01827932	0.04331407	0.04117554
Nigeria	Maize	0.00194973	0.00268457	0.01124211	0.03901889	0.04015009
Burkina Faso	Maize	0.00158174	0.00369686	0.02626520	0.03382954	0.03097056
Senegal	Maize	0.00106127	0.00281368	0.02088470	0.03361750	0.03085434
Guinea Bissau	Maize	0.00578081	0.01599571	0.08813826	0.04639926	0.05752965
Togo	Maize	0.00991938	0.02016226	0.09204900	0.10000162	0.08829238
Niger	Maize	0.00074933	0.00233211	0.01700366	0.03924410	0.02696511
Sierra Leone	Maize	0.00649442	0.01353270	0.06295945	0.05847860	0.07298443
Côte d’Ivoire	Maize	0.00105507	0.00227436	0.01596227	0.02108761	0.02094848
Benin	Cassava	0.00072762	0.00371670	0.03267188	0.01755675	0.01948215
Guinea	Cassava	0.00503696	0.00712383	0.02860466	0.06488639	0.06247591
Ghana	Cassava	0.00243144	0.00445531	0.02557120	0.04077162	0.04509468
Liberia	Cassava	0.00256122	0.00774713	0.04726972	0.03403242	0.03722366
Nigeria	Cassava	0.00185424	0.00277686	0.01209194	0.03993116	0.03968683
Togo	Cassava	0.00822577	0.01443074	0.06153987	0.07971147	0.07548504
Sierra Leone	Cassava	0.00185986	0.00511210	0.03064491	0.04245204	0.03845616
Côte d’Ivoire	Cassava	0.00112116	0.00263257	0.03064491	0.03194575	0.03279232
Benin	Rice	0.00071910	0.00496773	0.03950279	0.02117554	0.01622959
Guinea	Rice	0.00191753	0.00384247	0.02495658	0.04171192	0.04041576
Ghana	Rice	0.00088692	0.00214671	0.01569058	0.03193582	0.02447154
Liberia	Rice	0.00742250	0.02365423	0.10664631	0.06688038	0.05961841
Mauritania	Rice	0.01710714	0.04113854	0.09801760	0.04796407	0.07102053
Mali	Rice	0.00110347	0.00215285	0.01203484	0.03269070	0.02598336
Nigeria	Rice	0.00103784	0.00164344	0.00872079	0.03894496	0.03651457
Burkina Faso	Rice	0.00071322	0.00207747	0.01740650	0.03520919	0.02500710

Country	Crop	Minimum loss values	Mean loss values	Maximum loss values	RMSE on test set	RMSE on training set
Senegal	Rice	0.00146573	0.00401248	0.02403932	0.01891044	0.02720833
Guinea Bissau	Rice	0.01117795	0.02698722	0.12662517	0.08494826	0.08147735
Togo	Rice	0.00422155	0.00930080	0.05174400	0.04264234	0.05302349
Niger	Rice	0.00174473	0.00363644	0.01995699	0.03820201	0.03816552
Sierra Leone	Rice	0.00671424	0.01431605	0.07641059	0.07168690	0.07429794
Côte d'Ivoire	Rice	0.00259529	0.00527748	0.03147672	0.04233520	0.04016610
Libya	Wheat	0.00273258	0.00718171	0.04165258	0.04484544	0.04225213
Sudan	Wheat	0.00075528	0.00205568	0.01354778	0.03051089	0.02256988
Tunisia	Wheat	0.00919501	0.01323656	0.04416724	0.09916451	0.08879559
Morocco	Wheat	0.00685901	0.00816831	0.01921000	0.08206151	0.07935333
Egypt	Wheat	0.01897007	0.02592046	0.06760812	0.11164873	0.11744088
Algeria	Wheat	0.00295956	0.00403027	0.01379454	0.05042471	0.05085805
Kenya	Maize	0.00416994	0.00630878	0.02369906	0.05508986	0.05735548
Malawi	Maize	0.00751643	0.01245092	0.05270030	0.07838780	0.07839491
Zimbabwe	Maize	0.00495353	0.00621669	0.01702104	0.05888519	0.06400499
Mozambique	Maize	0.00064125	0.00131989	0.00947815	0.02158957	0.02111065
Ethiopia	Maize	0.00164969	0.00228679	0.00935273	0.03716214	0.03755907
Uganda	Maize	0.00255234	0.00430324	0.02320788	0.03919638	0.04631671
Tanzania	Maize	0.00142267	0.00214565	0.01043562	0.03307371	0.03440688
Zambia	Maize	0.00123491	0.00189236	0.00931647	0.03063931	0.03092630
Madagascar	Maize	0.00025017	0.00086813	0.00857775	0.01171361	0.01382347
Rwanda	Maize	0.00581005	0.01465933	0.06490663	0.05962805	0.05457481
Botswana	Maize	0.00195286	0.00471188	0.02337214	0.04084127	0.03649043
Eswatini	Maize	0.01283505	0.02411070	0.08909293	0.07191201	0.08209138
Namibia	Maize	0.00106955	0.00322694	0.02578572	0.02222601	0.02848192
Lesotho	Maize	0.01070529	0.02000675	0.06323701	0.07826008	0.09075452
South Africa	Maize	0.00158438	0.00195824	0.00549064	0.03786032	0.03599981
DRC	Cassava	0.00049400	0.00093600	0.00673400	0.01760200	0.01908400
Gabon	Cassava	0.00049300	0.00280600	0.02357000	0.00589000	0.00887100
Congo	Cassava	0.00136100	0.00328400	0.02398900	0.03479600	0.03287000
Central Africa	Cassava	0.00135600	0.00298900	0.01983500	0.02625900	0.02808400
Angola	Cassava	0.00031500	0.00072200	0.00545000	0.01251800	0.01243500
Cameroon	Cassava	0.00084300	0.00193800	0.01437900	0.02412500	0.03040200

Note: The root mean squared error (RMSE) on the test set was used for the model accuracy assessment.

Description of Preprocessing Steps

Data Preprocessing

The overarching goal of the data preprocessing procedure was to build the final dataset, which will be used to train the algorithm to learn the relationship between the input features such as the NDVI, LST-day, rainfall, and ET, with the targeted crop production values. This was performed by using available historical data from 2005, 2010, and 2017. The latter selected years are constraints dictated by the data available on the SPAM data portal. The entire process described below was completed with Spyder-Python 3.7.0 provided in the open-source individual Anaconda distribution.

Data Access

The first step of the preprocessing stage is accessing the relevant remote sensing products. For the input data such as NDVI, LST-day, and ET (which are MODIS data), the open-source Python pymodis package was used. The datasets come with a hierarchical data format, which consists of several scientific dataset (SDS) layers. However, for raster-type maps such as rainfall and production, TIF images were downloaded.

The Mosaicking Process

The methodology developed in this chapter embeds a mosaicking process that puts together different tiles from the same sensing date to cover a specific region of interest. Such a process is specific to satellite images due to their trajectory around Earth. The MODIS global sinusoidal tile grid is composed of 595 tiles, 460 of which are not filled. Tiles are 10 by 10 degrees at the equator with the following naming system: H_xx V_xx, where H_xx refers to the horizontal identification (ID), and V_xx, the vertical ID for any tile. The reference H_00 V_00 corresponds to the upper-left corner, and the lower-right corner is H_35 V_17. For our area of interest (the African continent), a dictionary of countries and their corresponding tiles was built for automation purposes. The open-source Python Geospatial Data Abstraction Library (GDAL) package was used for countries that require merging several tiles for complete geographical coverage.

Raster Extraction and Cleaning Process

This step aims to extract only the needed SDS layers from remote sensing products and to drop unreliable pixels. For NDVI, SDS layers 1 (NDVI data) and 12 (Pixel reliability) were used to extract NDVI layers and keep pixels that are labeled as good data (label 0) or marginal data (label 1). The exact process applies to the LST-day data where SDS layers 1 (LST data) and 2 (quality assurance data) were used. The process involves reading each raster of interest as a Python array with its corresponding index from the initial raster. Each data point that is not flagged as good or marginal data from the quality assurance layer is dropped. Then, a new raster is created using the open-source Python Rasterio package with the remaining data points.

Reprojection, Pixel Resampling, and Cropping

At this stage, the methodology dealt with three primary operations: reprojection, pixel resampling, and cropping. MODIS products that were selected for the predictive model were sinusoidal projected. For further computations with country administrative borders, both shapefiles and remote sensing products are required to have the same projection system. The GDAL package was also used to transform each raster projection system from sinusoidal to the 1984 world geodetic system (WGS84).

In addition, pixel size needs to be the same between RSPs and crop masks for further computations. The SPAM spatial resolution was chosen as a reference for other rasters. Therefore, an average resampling procedure using the GDAL package was used to aggregate smaller pixels from NDVI, LST-day, rainfall, and ET data to match the SPAM maps' pixel size. Finally, level 0 shapefile (national level) was applied to isolate the area of interest from the resulting map.

Crop Mask Application

Production maps retrieved from the SPAM portal were used to further isolate explanatory variables at areas where a specific crop is grown. A crop mask was built by allocating a value of one to each pixel with a production value greater than or equal to one and zero elsewhere. Therefore, by performing the arithmetic product of this mask with all the maps that were generated above, the result was new NDVI, LST-day, rainfall, and ET rasters at pixels where the selected crop is grown. However, for the 2020 dataset, we used the 2017 generated mask.

From Raster to Dataframe

Explanatory and response variables are required to build a supervised agricultural production model. In our case, each line (equivalent to a specific pixel) of the final dataset on which the model was built upon is a scenario. Therefore, the temporal resolution between inputs and outputs must match. However, production values are available at an annual basis, which means pixel values for input variables have to be annual, and for that, mean values were computed for each input feature during the crops' growing season only. For one crop, the final outputs would be 18 mean (or annual) rasters that are cropped to the region of interest and correspond to the 3 rasters (2005, 2010, and 2017) times 6 variables (5 as inputs and one as output). Each country would have the same number of scenarios as its number of pixels; most countries have tens of thousands of scenarios.