Generating Crop Distribution Maps for Uganda Using a Data Fusion Approach

Grace Cheng

Scarsdale High School

Scarsdale NY 10583

Abstract

This paper applies a Spatial Production Allocation Model (SPAM) created by researchers at the International Food Policy Research Institute (IFPRI) to the most recent national and sub-national agricultural production data of Uganda. By sequentially carrying out data disaggregation, optimization and allocation, gridded maps of Uganda's agricultural production for 28 crops were generated, which provides a more informative picture of the likely distribution of individual crops in the country that the national or sub-national level statistics alone could not reveal.

1. Introduction

Agriculture has provided livelihood and nourishment for thousands of years for the humanity. However, it is now a dominant force behind many environmental threats such as climate change, land and freshwater degradation, biodiversity loss (Foley et al. 2011, Rockstrom et al. 2009). In addition, the rising population and changing dietary patterns are estimated to double the food demand by 2050. Providing enough food for the seven billion people on the planet, while reducing the harmful effects of agricultural production is one of the biggest challenges in the 21st century (You, et al. 2014).

To face this challenge, we need tools, data, and information to provide more responsive and evidence-based analysis to improve the use and allocation of resources. The spatially explicit data on the scale and intensity of agricultural and pastoral systems is of essential importance. Data on agricultural production (i.e., harvested area, production quantity and yield) are usually representative at the national and sub-national administrative units, but this level of statistics does not give a sense of the diversity or spatial patterns in agricultural production, and is not spatially explicit. In the last decade, International Food Policy Research Institute (IFPRI) and its partners have been developing a spatial production allocation model (SPAM) for generating highly disaggregated, crop-specific production data. These include national or sub-national crop production statistics, satellite data on land cover, maps of irrigated areas, biophysical crop suitability assessments, population density, secondary data on irrigation and rainfed production systems, cropping intensity, and crop prices (Yu, et al. 2020). This information is compiled and integrated to generate the initial estimates of the spatial distribution of individual crops, which are then submitted to an optimization model using cross-entropy principles with area and production accounting constraints to simultaneously allocate crops into the individual "pixels" of a GIS database. The result for each pixel is the area and production of each crop produced, split by the shares grown under irrigated, high-input rainfed, low-input rainfed, subsistence conditions. The detailed spatial datasets fill in the data gap for exploring the social, economic, and environmental consequences of agricultural production at a highly disaggregated scale.

The original SPAM model has the estimates of crop area, yield, and production for 42 major crops under four farming systems across a global 5 arcmin grid. The current research attempts to utilize the SPAM model for a specific country – Uganda – by applying the recently collected national/sub-national level data, carrying out the optimization model, and finally generating the crop maps.

The paper is organized as follows: Section 2 provides background information of Uganda, focusing especially on agricultural production; Section 3 introduces the Spatial Production Allocation Model (SPAM); Section 4 describes data sources and applies the data to the SPAM model for Uganda; Section 5 shows the gridded maps for selected crops (plantain and rice) with interesting findings.

2. Background

Uganda is a landlocked country in East Africa. It has Kenya in the East, South Sudan in the North, Democratic Republic of the Congo in the West, Rwanda in the South-west, and Tanzania in the South. Agricultural land[1] in Uganda was reported at 71.89 % in 2018, according to World Bank collection of development indicators (Trade Economics, 2022). Agriculture is a key sector in this country's economy. In 2019 and 2020, 80% of the households engaged in agriculture, contributing to a total of 24% of its GDP. The government identifies agriculture as a driver of key growth opportunities with great potential to generate employment with positive multiplier effects on other sectors[2]. (Annual Agricultural Survey 2019 Report, Uganda Bureau of Statistics) Therefore, a more accurate spatially explicit data/statistics would help to better understand land management practices and provide evidence-based decision-making and policy development to improve the performance of the specific sector.

This paper focuses on applying the SPAM model to the most recent production data in Uganda. This practice produces gridded maps of Uganda's agricultural production at a 5 arcmin spatial resolution for 28 crops. The national level crop statistics are collected from FAOSTAT and the sub-national level distribution is from Uganda Bureau of Statistics.

The gridded maps are generated for each of the 28 crops produced in Uganda by harvested area and yield, providing the likely distribution of the production for each crop. This gives insights on land management practices in this country.

3. Model

The current research utilizes the Spatial Production Allocation Model (SPAM) first developed by You and Wood (2006). The main function of SPAM is to disaggregate crop statistics (e.g., harvested area, production quantity, and yield) by different farming systems and to further allocate such disaggregated statistics into spatially gridded units. The model uses a data fusion approach that

[1] Agricultural land refers to the share of land area that is arable, under permanent crops, and under permanent pastures.
[2] Includes manufacturing and services.

combines information from different sources and at different spatial scales by deploying various matching and calibration processes. Then, all the data elements are processed by the optimization model, which generates results at the grid level. Hence there are three submodules in a standardized SPAM model: disaggregation, optimization, and allocation.

3.1 Disaggregation

The first step for SPAM is to disaggregate crop statistics of agricultural production (e.g., the yield, harvested area, and total production) by administrative-unit levels (k), crop type (j), and farming system (l) from a coarser to finer scale.

For the administrative unit (ADM), three levels are considered – k = 0 (national level), 1 (subnational level 1), and 2 (subnational level 2) – and refer to the country specific administrative level as the statistical reporting units (SRUs; *SRU = k0, k1 or k2*). In this study, 28 crop types are considered *j = j1, j2, …, j28*. They are described in details in the below data section. Four farming systems are specified (*l = I, H, L, S*) representing the irrigated farming system (I), the rainfed high-input farming system (H), the rainfed low-input farming system (L), and the rainfed subsistence farming system (S). For details on these farming systems, please refer to the IFPRI research paper by You et al. 2014,

3.2 Optimization

A cross-entropy method is used to allocate the crop for each spatial grid (*i*). The cross-entropy builds on the idea of entropy from information theory (Shannon, 1948). Let *H* be the entropy of a discrete random variable *X*, which takes values in the alphabet $\aleph$ and is distributed according to $p\colon \aleph \to [0,1]$ such that $p(x) := P[X = x]$:

$$H(X) = E[I(X)] = E[\log p(X)]$$

where *E* is the expected value operator and *I* is the information content of *X*. The entropy can be written as:

$$H(X) = -\sum_{x \in \aleph} p(x) \log_b p(x)$$

where *b* is the base of the logarithm used, with possible values of 2, Euler's number *e* or 10 with the corresponding units of entropy the *bits* for *b* = 2, *nats* for *b* = *e*, and *bans* for *b* = 10.

The cross-entropy method used here tries to minimize the difference between two probability distributions for a given random variable. Specifically, the model iteratively minimizes the error between the pre-allocated shares of physical area (π_{ijl}) and the allocated shares of physical area (s_{ijl}) in each pixel *i* by crop *j* and production system *l*. The loss function is specified as follows:

$$\min_{\{s_{ijl}\}} CE(s_{ijl}, \pi_{ijl}) = \sum_i \sum_j \sum_l s_{ijl} ln s_{ijl} - \sum_i \sum_j \sum_l s_{ijl} ln \pi_{ijl}$$

where CE is the abbreviation for cross-entropy, which is defined as the log function of probability. The difference between $\{slns\}$ versus $\{sln\pi\}$ means the estimated probability and its prior probability π are minimized subject to certain constraints. The constraints for the optimization and the prior probability π are specified in more details in the technical paper by Wood-Sichra et al. (2016).

3.3 Allocation

The allocation module produces maps of area, yield and production for each grid *i* by crop *j* and farming system *l*, using the results of the optimization, i.e., shares s_{ijl}:

$$s_{ijl} = \frac{AllocA_{ijl}}{AdjCrop_{jl}},$$

Where $AdjCrop_{jl}$ is the total physical area of a given SRU for crop *j* at input level *l* to be allocated and $AllocA_{ijl}$ is the area allocated to grid *i* for crop *j* at input level *l*.

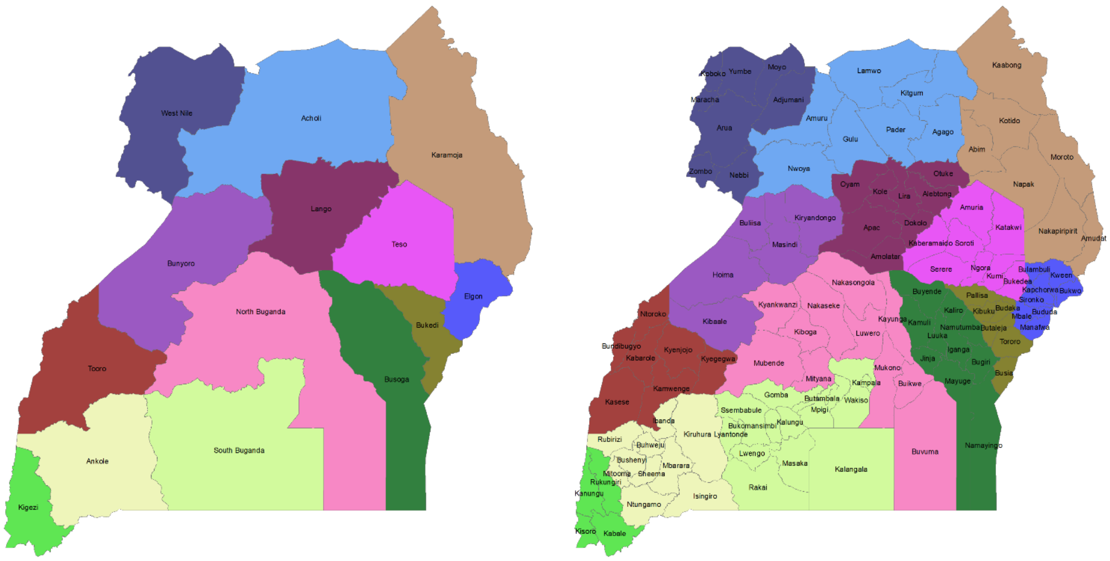
4. Data

Uganda's first level administrative units are "sub-region", and "districts" at the second level. There are 14 sub-regions in the country and 112 districts. Figure 1(a) shows the list of sub-regions in Uganda with each sub-region having its own color. Figure 1(b) displays the corresponding districts within each sub-region and the districts in the same sub-region distinguish themselves by the same color. The number of districts varies in each sub-region, ranging from 4 districts in one sub-region to 13 in one sub-region. This is the SRU for Uganda.

Data sources for crop statistics include FAOSTAT and Annual Agricultural Survey 2019 Report published by Uganda Bureau of Statistics. FAOSTAT provides the national level area and yield information for the 28 crops planted in Uganda. Table 2 lists the most recent 3-year average[3] of the area and yield for these 28 crops.

[3] 2018-2020 data is used to calculate the 3-year average.

Figure 1: List of sub-regions and districts within Uganda



(a) Sub-regions

(b) Districts

Annual Agricultural Survey 2019 Report provides the sub-region level area and yield statistics for each crop. However, a few crops (such as wheat) only have national, not sub-regional, level data. For these kind of crops, sub-regional level statistics are replaced with a value of *-999*, representing missing information, in the model run. Table 3 gives an example of the sub-regional area and yield information.

The SPAM model takes the national level statistics and then re-assigned the national level area or yield information to the sub-regional level based on the sub-regional distribution.

Table 2: 2018-2020 average of area and yield for the crops planted in Uganda

	area (ha)	**yield (hg/ha)**
bean	432,157	16,769
cassava	1,557,814	33,340
chickpea	8,524	6,126
cocoa	72,025	4,860
coffee	538,988	5,490
cotton	88,000	13,558
cowpea	35,221	3,715
groundnut	312,000	8,793
maize	1,198,807	27,237
onion	80,453	39,519
other cereal (millet)	130,761	13,967
other oil	51,851	8,976
other tropical fruit	411	84,773
other vegetable	152,290	67,308
pigeon pea	44,921	3,544
plantain	1,512,467	42,401
potato	51,112	63,010
rice	76,594	28,484
sesame seed	213,333	6,781
sorghum	286,280	8,513
soybean	49,667	20,246
sugarcane	79,454	703,967
sunflower	270,667	9,941
sweet potato	357,489	42,009
tea	30,965	21,320
tobacco	22,421	14,351
tomato	6,454	57,616
wheat	15,527	15,504

Table 3: Sub-regional area and yield statistics for crop "bean"

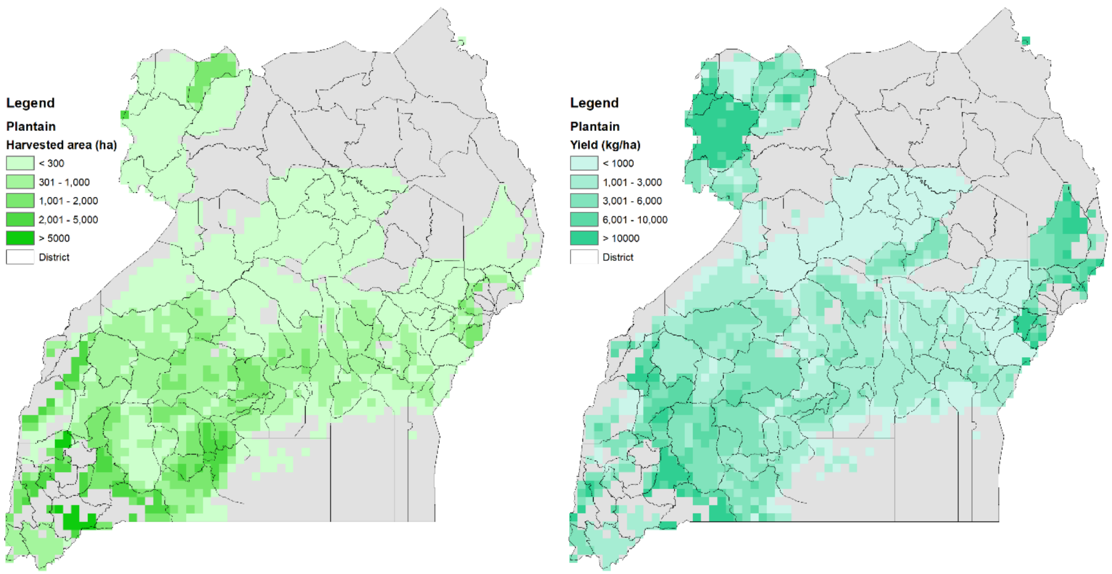
	bean	
	area (ha)	yield (kg/ha)
S. Buganda	97,738	600
N. Buganda	123,709	600
West Nile	17,915	600
Lango	63,081	400
Acholi	12,805	400
Kigezi	34,800	600
Bunyoro	117,347	700
Tooro	89,453	600
Busoga	19,690	400
Teso	7,738	300
Bukedi	9,471	300
Elgon	72,970	400
Karamoja	4,102	200
Ankole	94,031	800

5. Discussion

In this section, the results from the SPAM model are displayed for the selected crops in Uganda. The model makes plausible estimates of the crop distribution within disaggregated units. The outputs are shown as overlays on the map indicating how each specific crop is spatially allocated in the country.

Figure 2 presents the maps with crop masks of the harvested area and yield for plantain. Plantain is one of the most important crops in Uganda as it is among the top 10 banana-producing countries in the world. According to the maps, plantain is mainly produced in southern Uganda, with the most intensive planting in the southwest. Yield of this crop varies across the country. On average, the southern area has higher yield comparing to the other parts of the country. But the most productive areas appear to be located in the northwest followed by some areas in the east.

Figure 2: Harvested area and yield maps for Plantain in Uganda

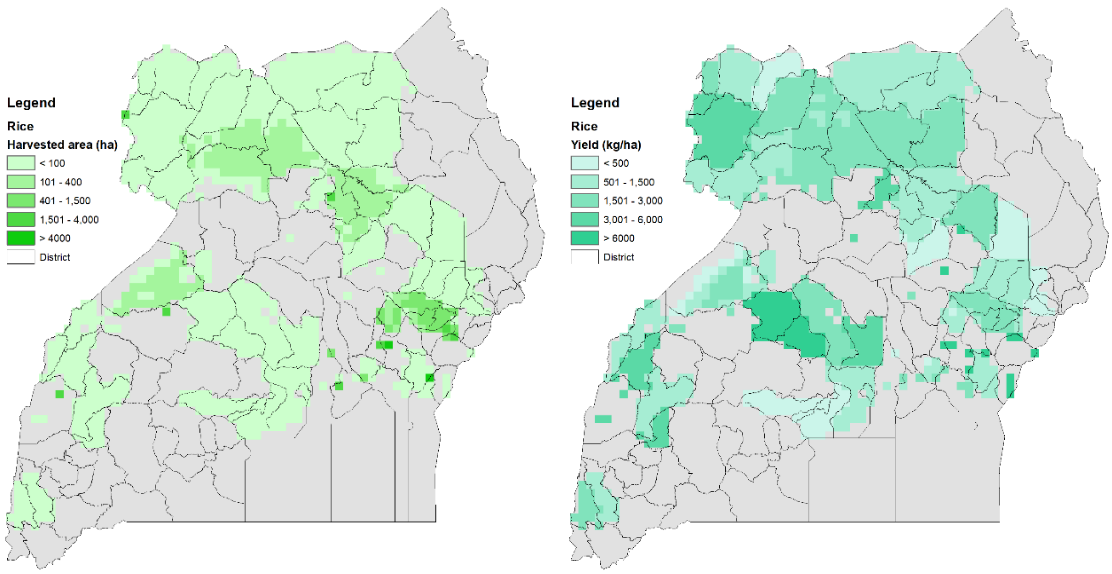


(a) Harvested area (b) Yield

Figure 3 demonstrates the maps with crop masks of the harvested area and yield for Rice. In Uganda, rice is considered a strategic crop with the potential to contribute to increasing rural incomes. This crop's production is on an upward trend in the country. According to the map, major rice production is in a small area in eastern Uganda followed by certain areas in western Uganda. Compared to plantain, it is obvious to see that the harvested area of rice is much less widely spread as rice is not a major crop in this country. Yield is higher in the center of the country and in certain areas of the northern part.

In general, these maps provide additional insights on the supply of agricultural products in Uganda which is in increasing demand with good economic value but also highly sensitive to climate change. The maps can provide the basis for more precise acreage and yield analysis, incorporating variables including weather patterns, soil conditions, vegetative health as well as input usage. The information could be used to better target agricultural development policies, increasing food security and growth with minimal environmental impacts. (Yu, et al., 2020).

Figure 3: Harvested area and yield maps for Rice in Uganda



(a) Harvested area (b) Yield

References

Food and Agriculture Organization of the United Nations (FAOSTAT). 2022. https://www.fao.org/faostat/en/#home (accessed August, 2022)

Foley, J. et al. 2011. Solutions for a cultivated planet. Nature 478, 337–342. doi:10.1038/nature10452

Rockstrom, J. et al. 2009. A safe operating space for humanity. Nature 461, 472–475.

Shannon, Claude E. (1948). "A Mathematical Theory of Communication". Bell System Technical Journal. 27 (3): 379–423

Trade Economics, 2022. https://tradingeconomics.com/uganda/agricultural-land-percent-of-land-area-wb-data.html

Portmann, F. T., S. Siebert, and P. Döll (2010). MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Global Biogeochemical Cycles, 24(1), n/a–n/a. doi:10.1029/2008GB003435

You, L. and S. Wood (2006). An entropy approach to spatial disaggregation of agricultural production. Agricultural Systems, 90(1-3), 329–347. doi:10.1016/j.agsy.2006.01.008

You, L., S. Wood, and U. Wood-Sichra (2009). Generating plausible crop distribution maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach. Agricultural Systems, 99(2-3), 126–140. doi:10.1016/j.agsy.2008.11.003

You, L., S. Wood, U. Wood-Sichra and W. Wu. (2014). Generating global crop distribution maps: From census to grid. Agricultural System, DOI: 10.1016/j.agsy.2014.01.002

Yu, Q., You, L., Wood, S. et al. (2020). A cultivated planet in 2010 – Part 2: The global gridded agricultural-production maps. Earth System Science Data, December 2020, page 3545 – 3572.

Wood-Sichra, U., A. Joglekar, and L. You (2016). Spatial Production Allocation Model (SPAM) 2005: Technical Documentation. Working Paper, International Food Policy Research Institute, Washington DC.