Optimizing Ethiopia's Cereal Crop Storage Infrastructure: A Geospatial and Network Science Approach

Isabella Maria Rangel

April 11, 2025

Acknowledgements

I would like to express my sincere gratitude to my thesis advisor, Professor William Grieser, for his thoughtful feedback and enthusiasm throughout this project. I also thank Sam Armon for his support during the Directed Reading Program in network science.

Contents

List of Figures

Chapter 1: Introduction

1.1 Thesis Statement

This study examines how strategically optimizing the locations of cereal crop storage facilities in Ethiopia can potentially improve supply chain efficiency, reduce postharvest losses, and enhance resilience to disruptions. Using a combination of network science and geospatial optimization techniques, this study identifies optimal storage facility placements that strengthen the agricultural supply chain.

1.2 Motivation

Africa contains 25 percent of the world's arable land yet produces only 10 percent of global agricultural output (International Fund for Agricultural Development (IFAD), 2025). Moreover, Africa remains a net importer of agricultural goods, despite its substantial production potential. Figure 1.1 shows aggregate cereal yields per unit area for five of the seven world continents, underscoring Africa's comparatively low productivity. This disparity highlights the urgent need to improve efficiency in the continent's agricultural sector. According to a recent study by Goedde et al. (2024), Africa could produce an additional 2.6 billion tons of cereals and grains—enough to increase global supply by 20 percent. However, to realize this potential, the study estimates that the continent must invest approximately $8 billion in storage infrastructure.

Africa's Agricultural inefficiency is perhaps most acute in Ethiopia, The continent's largest cereal producer as of 2022 (Sasu, 2022). In 2023, agriculture accounted for 32 percent of Ethiopia's GDP (African Development Bank, 2024) and 80 percent of its export

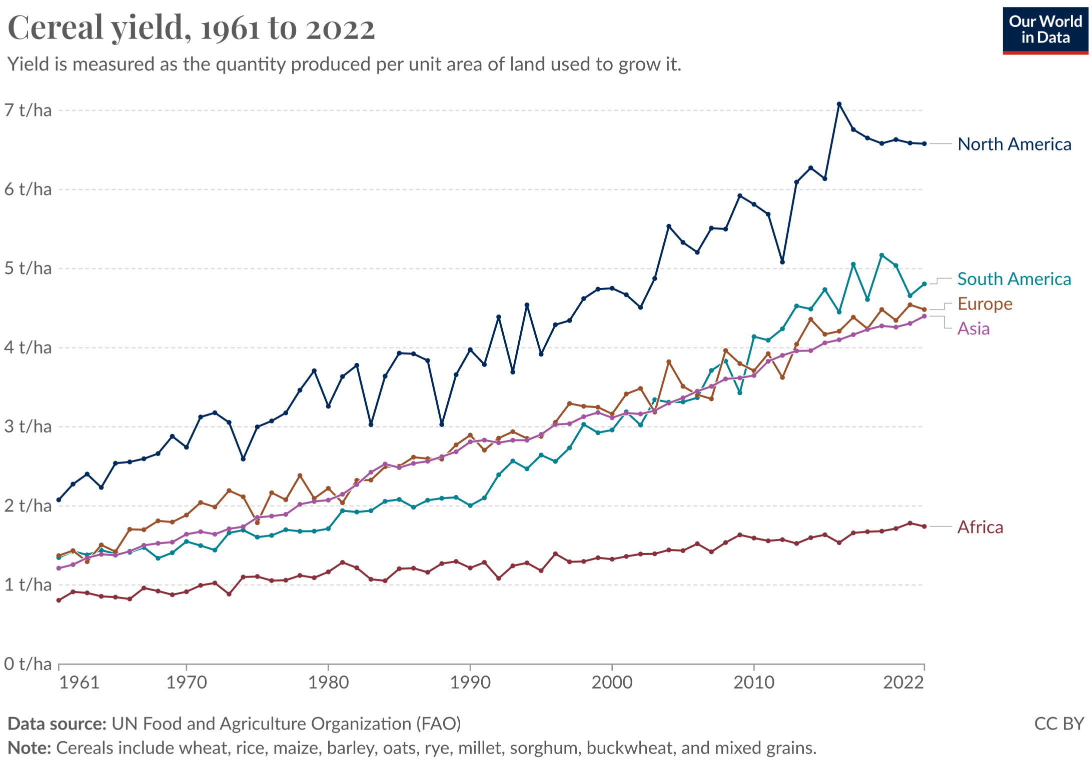


Figure 1.1: Cereal yield comparison by continent
Image Sourced From: Ritchie et al., 2023

revenue, with approximately 99 percent of those exports consisting of unprocessed commodities (Strubenhoff, 2021). Beyond its economic significance, agriculture is vital for food security: cereals alone provide 70 percent of Ethiopia's caloric intake (Berhane et al., 2012). Despite this, an estimated 15.8 million people required emergency food assistance in 2024, and more than half of children under five were malnourished (World Food Programme, 2025). In 2022, 19.7 percent of the population was food insecure, even though agriculture employed 63 percent of the country's labor force (World Bank, n.d.-a, n.d.-b).

The crisis has since deepened. Ethiopia is now facing famine conditions in its northern Tigray region, which was embroiled in conflict with the central government from 2020 to 2022. This conflict led to lost harvests, mass displacement, and widespread destruction of infrastructure. Climate shocks compounded the devastation: a historic drought from 2020–2022 and catastrophic floods triggered by El Niño in 2023 further battered the region

(Goddard, 2024). As of the most recent assessments, 91 percent of Tigray is at risk of "starvation and death" (Yibeltal, 2023).

A major contributor to Ethiopia's agricultural challenges is the frequent and substantial postharvest loss of cereal crops. These losses amount to 10 percent of the country's annual budget and represent enough food to feed 23 million people each year. Research points to weak infrastructure and insufficient institutional support as key drivers of these losses (Food and Agriculture Organization of the United Nations (FAO), 2017). In particular, the lack of adequate storage facilities and reliable transportation networks severely disrupts the movement of goods from farms to markets (International Development Research Centre (IDRC), 2025).

Natural disasters further exacerbate the problem by damaging existing infrastructure and causing prolonged delays. For example, in 2024, heavy rainfall triggered landslides in Ethiopia's Gofa Zone, destroying roads and cutting off access to the region (Li et al., 2025). Many rural roads become impassable during the rainy season, forcing farmers to sell their harvests early—often at lower prices—before the rains begin (Bank, 2018). Transport inefficiencies are costly: an additional hour of travel increases cereal transportation costs by an estimated 20 percent (Minten et al., 2012).

Proper postharvest management has the potential to substantially reduce crop losses, according to Teferra, 2022. One promising approach involves introducing big data analytics to support systematic decision-making and enhance efficiency, productivity, and food security Hassen and Chen, 2022. In particular, optimizing the placement of storage facilities using data-driven methods could dramatically improve Ethiopia's agricultural supply chain resilience and reduce postharvest losses. However, doing so requires overcoming major challenges, including inadequate transportation infrastructure, insufficient storage capacity, and limited market access for farmers and consumers.

The African Postharvest Losses Information System (APHLIS) tracks dry weight, nutritional, and financial losses across supply chain levels, primarily focusing on cereals in

Sub-Saharan African countries. The information illustrates the importance of final crop storage optimization. In 2019, Ethiopia lost over 8 million USD (1 percent) of teff during transport and 21 million USD (2.7 percent) in market storage (African Postharvest Losses Information System (APHLIS), 2023). Improving infrastructure for crop storage would not only improve market storage. If this system becomes more efficient, that would mean making it easier for more farms to store their harvested cereals outside of homes. Household-level storage contributes to a large portion of postharvest losses. Figure 1.2 shows the postharvest losses for Ethiopian cereal crops at each stage of the value chain in USD for 2019.

Postharvest losses occur at multiple points in the agricultural value chain after harvesting, encompassing both quantitative and qualitative degradation. Key contributors include pest infestations, poor storage conditions, transportation delays, and improper handling. As Kumar and Kalita emphasize, minimizing cereal losses is a cost-effective strategy for enhancing food security. Reducing these losses not only helps alleviate hunger but also improves farmers' financial outcomes—an especially critical concern in Sub-Saharan Africa, where maize alone accounts for over 35 percent of caloric intake (Kumar & Kalita, 2017). Indeed, International Fund for Agricultural Development (IFAD), 2025 estimate that growing Ethiopia's agriculture is 11 times more effective in reducing extreme poverty than growth in other sectors.

Improper produce storage contributes greatly to these losses. An Ethiopia case study found that 46 percent of farmers stored these harvested cereals inside their homes. 39 percent stored their harvested crops in traditional "gotera." Other common storage options include undergound pits and polypropylene bags. Only one percent stored their crops in metal silos. Because of these storage practices, crops become more vulnerable to pests. Researchers found that pests caused losses ranging from 9 to 64.5 percent of maize production and 13 to 95 percent of sorghum production(Berhe et al., 2022). A study on minimizing Ethiopian postharvest losses in teff found that storing cereals in metal silos "significantly decrease losses during storage, mitigating both biotic and abiotic factors" (Tiguh et al., 2024). While pests and bacteria are biotic factors mentioned previously, abiotic risk factors such as moisture or fire are also mitigated with storage in metal silos.

While storing crops in metal silos prevents post-harvest losses, the vast majority of farmers in Ethiopia are not using this method. Smallholder farmers produce the majority of Ethiopia's agricultural output - 90 percent. (Haile et al., 2022a) Researchers in the Gubalafto District of northeast Ethiopia studied farmers' willingness to pay for metal silos. Farms with access to the market are 16 percent more likely to pay for these silos than farms without access. Farms with land ownership are 10 percent more likely to be willing to pay. Literacy had the greatest influence on farmers' willingness to pay. Households with higher educational status were 20 percent more likely to be willing to pay. (Teshome et al., 2023)

Taken together, these challenges illustrate a critical bottleneck in Ethiopia's agricultural system: a lack of effective, accessible, and data-informed storage infrastructure. Despite promising technologies like metal silos and growing evidence on their effectiveness, adoption remains low due to infrastructural, institutional, and behavioral constraints. By applying geospatial optimization and network analysis to identify strategic storage locations, this study offers a scalable framework for reducing postharvest losses and improving food security. Ethiopia's case highlights broader lessons for agricultural resilience across Sub-Saharan Africa.

	All steps	Harvesting/field drying	Further drying	Threshing and Shelling	Winnowing	Transport from field	Household-level storage	Transport to market	Market storage
Maize	392,965,567	126,288,802	73,633,060	23,326,953	-	41,317,968	120,008,512	3,252,597	5,137,674
Rice	-	-	-	-	-	-	-	-	-
Sorghum	156,839,752	57,032,339	-	43,071,324	-	24,989,344	27,398,217	1,200,090	3,148,437
Millet	-	-	-	-	-	-	-	-	-
Wheat	352,887,106	110,574,611	-	83,425,634	-	57,504,098	83,807,941	4,850,234	12,724,589
Barley	-	-	-	-	-	-	-	-	-
Fonio	-	-	-	-	-	-	-	-	-
Oats	-	-	-	-	-	-	-	-	-
Teff	404,451,340	112,498,422	-	108,560,977	74,829,531	72,958,792	5,998,690	8,170,257	21,434,670

Figure 1.2: Ethiopia post-harvest losses at each stage of the value chain in USD for 2019
Image Sourced From: African Postharvest Losses Information System (APHLIS), 2023

1.3 Related Literature

The analysis in this study draws from the literature on geospatial optimization, which identifies the best possible solution for geographic problems by maximizing or minimizing an objective function. An example of this is planning bus stops aiming to maximize ridership with spatial constraints prevent them from placing stops everywhere (Warf, 2010). This is a classic facility location problem (FLP). Introducing constraints, such as facility capacities, transforms a standard FLP into a capacitated facility location problem, which researchers can formulate as a mixed-integer programming problem (Alenezy, 2021, p. 1088).

Researchers use mixed-integer programming to solve a wide range of facility location problems, including wastewater treatment placement, emergency transfer locations, urban telecommunication architecture, and railway maintenance facility placement (Chandra et al., 2021; D'Andreagiovanni et al., 2016; Emami et al., 2024; Tönissen et al., 2019). The problem of identifying cereal crop storage locations to limit transportation costs, while adhering to real-world constraints aligns with this structure.

Optimization models are essential for effectively identifying optimal locations within agricultural supply chains by systematically weighing multiple objectives, such as minimizing total transportation and operational costs, while also accounting for spatial distribution, infrastructure, and production. Neto et al. apply a multi-objective MIP approach to a territory partitioning problem in Paraná, Brazil, determining where to construct new grain silos by minimizing regional transportation costs and storage imbalances (Neto et al., 2017). Their model improved logistical efficiency by clustering municipalities with aligned production and storage needs. In another research study, Mogale et al. develop a multi-objective, multi-modal, multi-period mixed-integer non-linear programming (MINLP) model to optimize silo placement across four supply chain areas in India. This study incorporated a non-linear dwell time component representing logistical delays, minimizing both total supply chain cost and lead time. This model effectively reduced storage delays and grain losses, improving government infrastructure planning (Mogale et al., 2018). These studies highlight the practical impact of MIP-based frameworks in improving the efficiency and resilience of grain storage networks.

Building on this foundation, Mogale et al. extend the application by incorporating sustainability and risk into the objective. Their model minimizes not only transportation and storage costs but also post-harvest losses, carbon emissions, and supply chain risk through a MIP model. This models five levels of the grain supply chain and implementing an enhanced particle swarm optimization algorithm, they demonstrate a more holistic and scalable approach to storage facility planning in food systems (Mogale et al., 2020).

The analysis in this thesis also draws heavily from network science. A network is a system of interconnected elements. Networks can have various structures and attributes. Networks consist of nodes, connected by edges. These links can be directed, weighted, and have multiplicity. These edges should model real-world characteristics. For example, some edges can represent transport networks with differing flow rates and capacities.

Researchers have applied network science to test and enhance resilience in supply chain networks. Percolation theory is the primary method for testing resilience. For instance, removing individual nodes or edges of the network, one element at a time, is an effective approach for analyzing the resiliency of the system to different failures or attacks. (Barabási, 2016) Insights from these studies can be applied to help optimize the supply chain network for cereal crops in Ethiopia.

This method can be added as another factor to the facility location optimization problem. Studies demonstrate that percolation-based analyses can provide actionable insights into reroute costs resulting from facility failures. In a large-scale study of the U.S. medical equipment supply chain, researchers modeled random failures and targeted attacks to identify how disruptions to highly connected suppliers drastically increased rerouting complexity and cost (Lavassani et al., 2023). These findings highlight the value of identifying central nodes whose failure would fragment the network and force costly detours. Similarly, Snyder and

Daskin introduce a facility location model that explicitly incorporates expected failure costs by assigning customers to a hierarchy of backup facilities, each with associated transportation costs and failure probabilities (Snyder & Daskin, 2005). Their model quantifies rerouting cost under disruption scenarios, making it possible to prioritize facility placements that minimize both primary and contingency costs. These approaches can be adapted to Ethiopia's cereal supply chain to improve network resilience and reduce the financial impact of postharvest distribution failures.

In a paper titled "Structural Measures of Resilience for Supply Chains" introduces two models: homogeneous and heterogeneous. This differentiation looks at numbers of suppliers. The heterogeneous model is what most correlates to the structure of this network. This is because supplier sizes and vulnerabilities vary by region. Not all areas produce the same yield. This model "assumes that failures can be correlated with each other" (Papachristou & M, 2024). In the case of the cereal crop supply chain in Ethiopia, failures are often correlated due similar climate and environmental factors. This can be helpful when looking at storage or transportation failures due to major climate events.

Chapter 2: Methodology

This section proposes an optimization model to identify final cereal crop storage facility locations, which minimize the total transportation cost from farm areas to storage facilities. After proposing a basic model, the analysis systematically introduces varying levels of complexity. Each added constraint or layer to the objective function can simulate the structure of the real-world scenario. For example, additional constraints can account for limited marketing rates (where only a fraction of cereals make it to market), budget constraints on facility construction, and limited storage facility capacity.

2.1 Data Sources

The analysis in this thesis relies on diverse data sources to address the research questions and optimize Ethiopia's cereal crop storage locations. Table 2.1 summarizes the key information gathered from each source, the resolution of data, and its purpose.

Dataset	Resolution	Key Information	Purpose
SPAM	10 x 10 km	Production, location coordinates	Identifies crop production per gridcell of area.
FAO	1 x 1 km	Suggested storage location coordinates	Identifies suggested final cereal crop storage locations.
GRIP	8 x 8 km	Road type, surface type, length, location	Data to calculate transport distance between farm and storage locations.

Table 2.1: Summary of Datasets

2.1.1 Spatial Production Allocation Model

The Spatial Production Allocation Model (SPAM) provides 10 x 10 km grid-cell resolution data with four key indicators: physical area, harvest area, production, and yield for 46 crops. These are further split by rainfed, irrigated, and combined production systems, resulting in approximately 500 million records for the year 2020. (International Food Policy Research Institute (IFPRI), 2024) SPAM uses cross-entropy optimization to allocate production to spatial grid cells. This method ensures that the results are statistically plausible, given the available data.

Figure 2.1 illustrates the process of data collection and optimization. SPAM begins with production information provided by administrative regions. These can be at the country-level, but smaller administrative region data results in higher accuracy. This step is identified as (a) in the figure. Pre-processing is done on this administrative-level data. Then, SPAM uses land cover data, which identifies area by land cover classes base on satellite imagery. Examples of some land cover classes include urban, forest, water, and grassland. Crop suitability is considered using agroecological data. Additional information on available data fields for the SPAM dataset can be found in the Appendix.

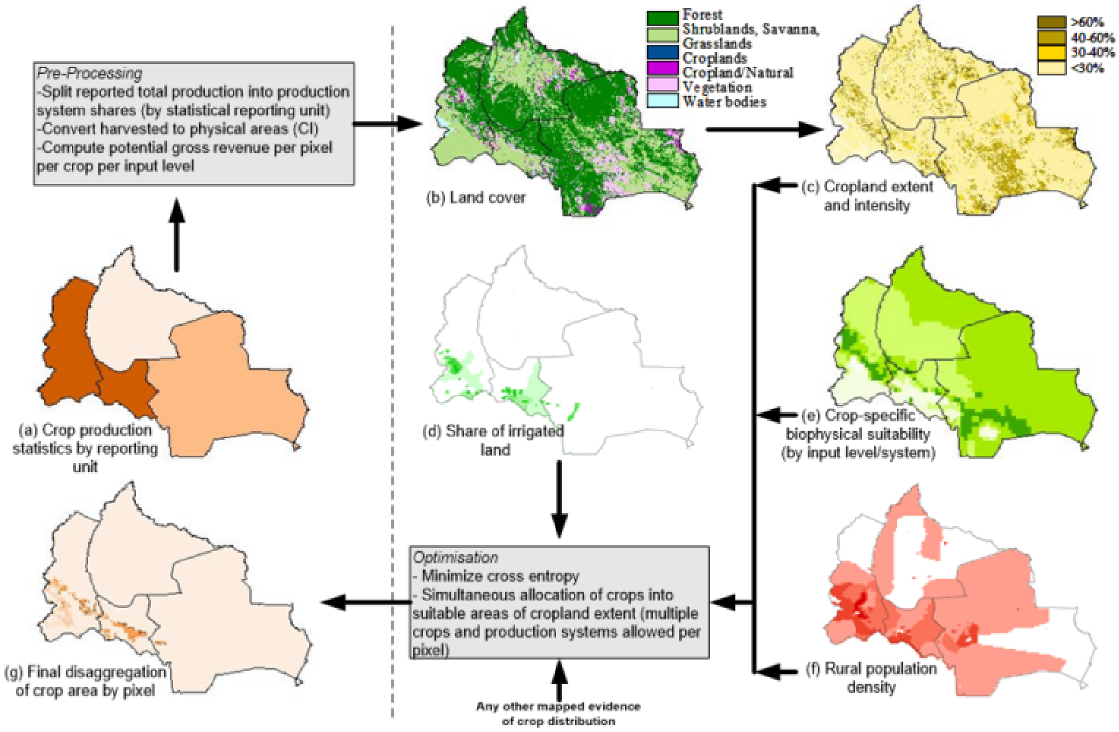


Figure 2.1: Spatial Production Allocation Model Process
Image Sourced From: SPAM, 2025

The SPAM dataset accounts for nearly every country in the world. However, there are some gaps for nations like: Liechtenstein, Palau, and Vatican City, and South Sudan. Small island nations are disproportionately excluded. Figure 2.2 shows the global map of production data available.

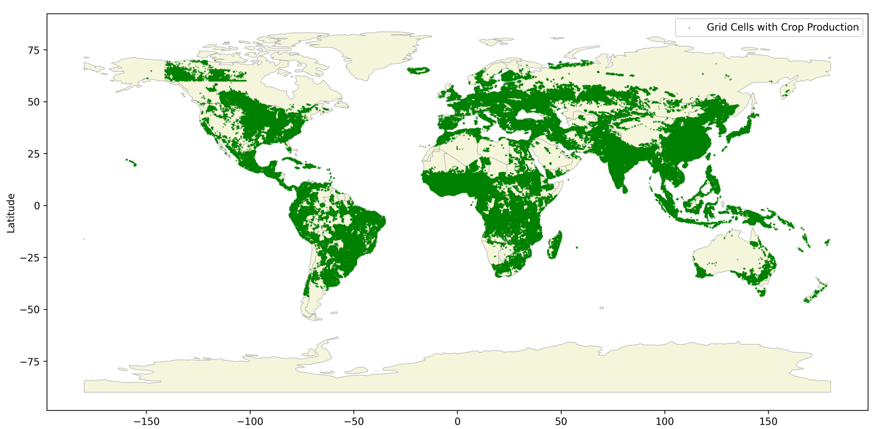


Figure 2.2: SPAM Global Crop Production Map
Data Source: International Food Policy Research Institute (IFPRI), 2024

As noted, this study focuses on agriculture in Ethiopia. After filtering the data for Ethiopia, the dominance of cereal crops becomes apparent. Maize is the most-produced crop in Ethiopia by tonnage, followed by 4 other cereal crops. The distribution of total production for cereal crops by administrative region be visualized as a heatmap. Most production happens in the ADM1 regions of Amhara and Oromia. There is a concentration of production around Lake Tana in the north. The lake is the white gap in the map.

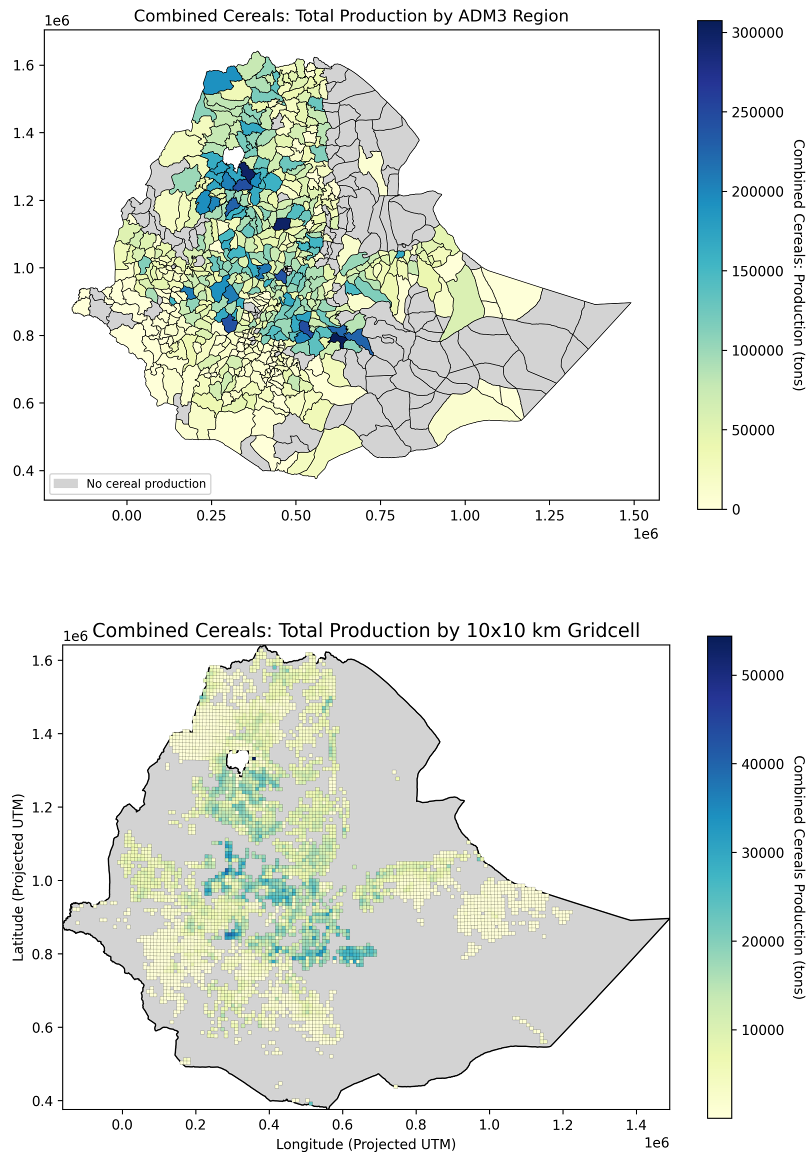


Figure 2.3: Production by region.
Data Source: International Food Policy Research Institute (IFPRI), 2024; Runfola et al., 2020

Figure 2.4 shows a mapping of the farm areas, color coded by which crop dominates production in tons for each 10x10 km gridcell of area. Dominance is determined by identifying the crop with the highest total production (in tons) within each farmland grid cell.

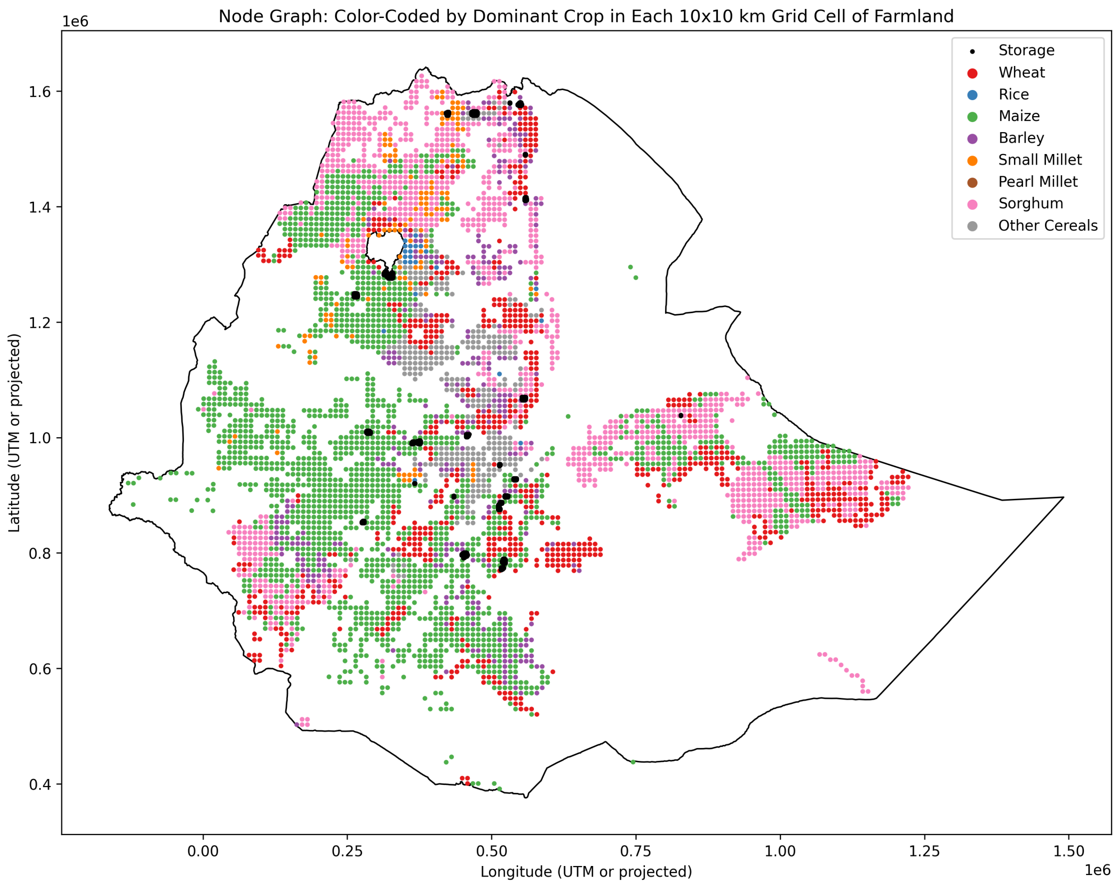


Figure 2.4: Farm nodes color-coded by dominant cereal crop produced.
Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024

The dominance ratio for each gridcell is the fraction of total production contributed by the dominant crop. The dominant crop is the crop with the highest production in that grid cell. Figure 2.5 displays the crop dominance ratio distribution for each 10x10 km gridcell of area. Most area gridcells have a dominance ratio between 0.35 and 0.65, as displayed by Figure 2.5. There is a noticeable decrease in the quantity of farm gridcells with a dominance ratio between 0.65 and 0.95. Then, there is a peak at 1.0. Additionally, there is high crop diversity within each 10x10 km gridcell. This makes sense because 90 to 95 percent of agricultural production in Ethiopia comes from smallholder farms. (International Fund for Agricultural Development (IFAD), n.d.).

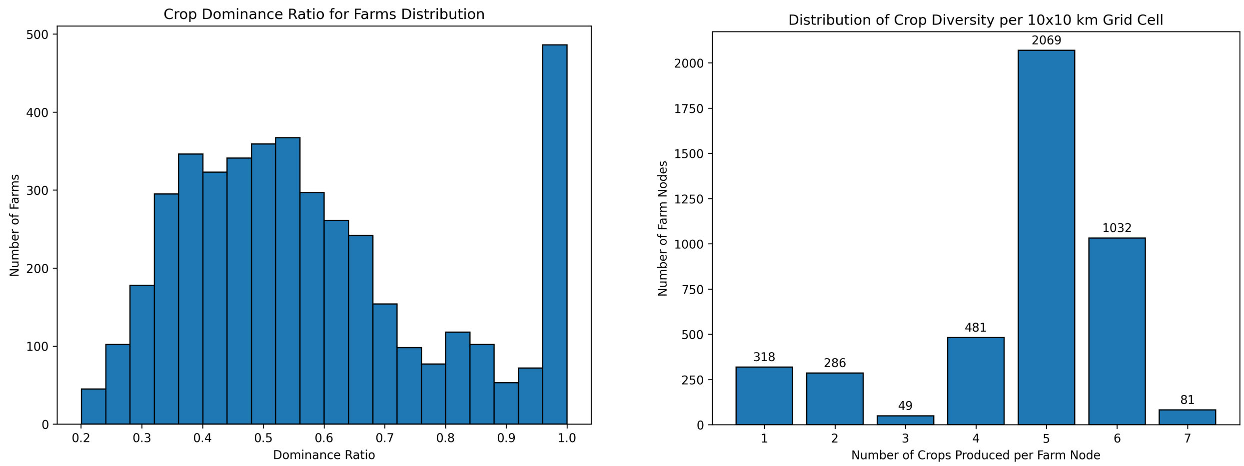


Figure 2.5: Crop dominance and diversity distributions for each 10×10 gridcell.
Data Source: International Food Policy Research Institute (IFPRI), 2024

Table 2.2: Count of 10x10 Gridcells Producing Individual Crops (and Dominance Counts)

Crop	Any Production	Primary Dominant	Secondary Dominant
Maize	4043	1965	1129
Sorghum	3729	1125	723
Other Cereals	3664	318	1100
Barley	3656	320	492
Wheat	3450	432	522
Small Millet	1283	134	213
Rice	240	22	137

Data Source: International Food Policy Research Institute (IFPRI), 2024

2.1.2 Food and Agriculture Organization of the United Nations

Data on global crop yields, trade flows, and agricultural productivity come from the Food and Agriculture Organization (FAO) of the United Nations. In 2021, the FAO published the data set: "Crop Storage Final Location: Cereal (Ethiopia - ~ 1Km)" as part of their Hand-in-Hand initiative, which aims to "accelerate the agricultural transformation and sustainable rural development to eradicate poverty and end hunger and all forms of malnutrition" (AmeriGEOSS, 2025).

The factors used to identify these top locations were access to financial resources, proximity to transport, broadband connectivity, and other infrastructure factors. Potential areas are identified in 1 km grid-cells. This means that for a suggested location larger than 1 km, many grid cells will be identified. This grid size is much smaller than the 10 km grid-cells used for the farm nodes. The FAO identify 853 1x1 km gridcells as potential optimal storage locations. These gridcells are clustered into 25 contiguous regional areas rather than 853 distinct storage locations. Thus, without much loss of generality, collapsing the grid cells into 25 will simplify the analysis.

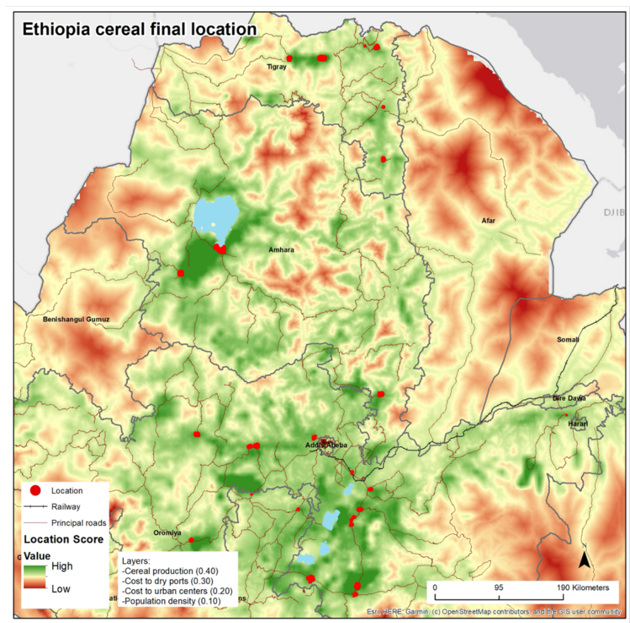


Figure 2.6: FAO location scores for Ethiopian cereal crop final storage locations.
Image Sourced From: Food and Agriculture Organization of the United Nations, 2022

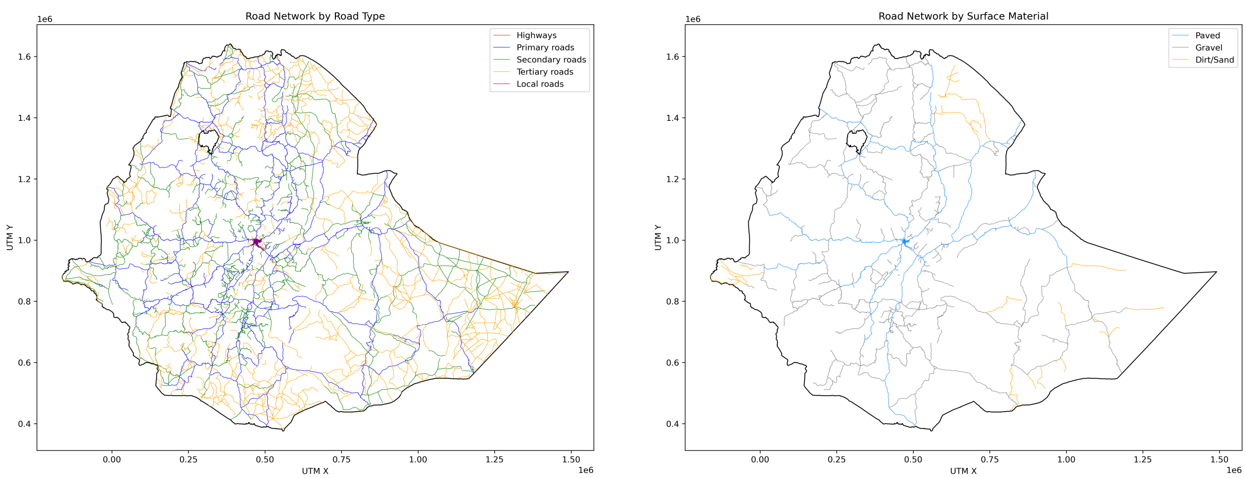
Table 2.3: Summary Statistics for Recommended Storage Areas

Statistic	Value
Number of recommended areas (after merging)	25
Minimum area size	2 km^2
Median area size	28 km^2
Mean area size	34.12 km^2
Maximum area size	130 km^2

Data Source: Food and Agriculture Organization of the United Nations, 2022

2.1.3 Global Roads Inventory Project

Information on road type, surface type, length, and geographic coordinates come from the Global Roads Inventory Project (GRIP). These data are standardized to align with the United Nations Spatial Data Infrastructure (UNSDI) transportation data model. GRIP uses global and regional sources to create vector and raster datasets, at a resolution of 8 x 8 km grid-cells. This means that roads within 8 km of each other will be aggregated and lose detail. While this resolution is not suitable for precise navigation, it is a reasonable option to calculate cost of transportation between farms and storage facilities in Ethiopia, considering its detail in road type and surface material (Meijer et al., 2018).



(a) Road types.

(b) Road surface material.

Figure 2.7: GRIP Road Attributes
Data Source: Meijer et al., 2018

Many "local roads" in the urban area of Addis Ababa are disconnected from eachother, based on observable data. Further investigation confirms there are other road types missing that allow additional transport through the region. The final data exclude this road type for the purposes of analysis. After these adjustments, there remain many disconnected segments in the network, as displayed in Figure 2.9.

I fixed this by looking at the road nodes with degree 1, which are endpoints. The resolution of this dataset is 8km, so if two endpoints were within 10km of eachother, but identified as different segments, they were merged together. There was also a clear issue with the primary roads in the north. There was a larger disconnected segment, which should have connected a major road. This had to be manually adjusted. Figure 2.8 visualizes the manually added segments. Finally, to ensure that the network was fully connected for valid routing computations, any remaining isolated road edges were removed from the dataset. Figure 2.9 visualizes these removed segments.

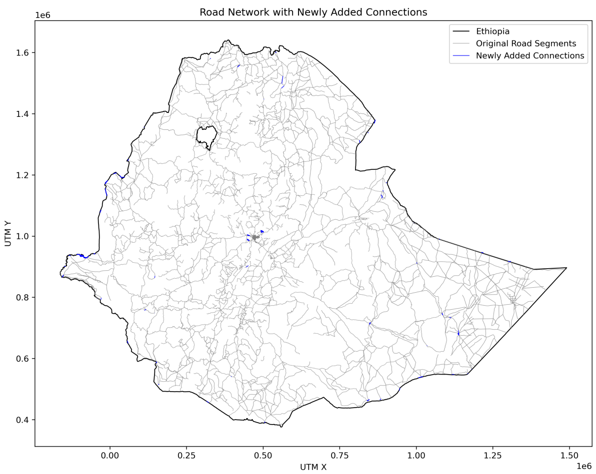


Figure 2.8: Segments Added

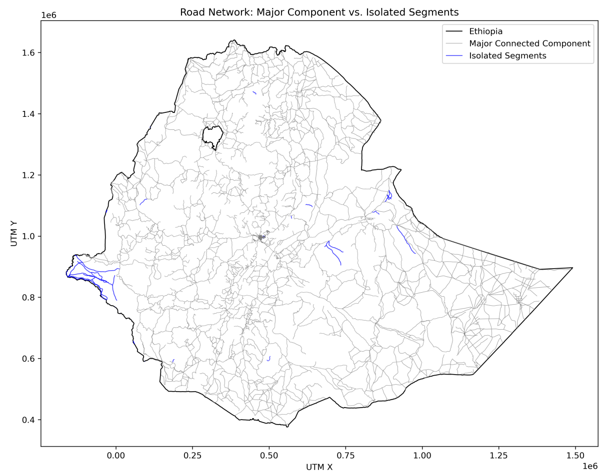


Figure 2.9: Isolated Road Segments

Data Source: International Food Policy Research Institute (IFPRI), 2024

2.1.4 Cleaned Data

Figure 2.10 presents the three datasets after processing and cleaning. The black lines represent the road network in Ethiopia. The green squares represent the gridcells of producing farmland, identified by the Spatial Production Allocation Model. The red triangles represent the gridcells identified as suggested crop storage locations by the Food and Agriculture Organization of the United Nations.

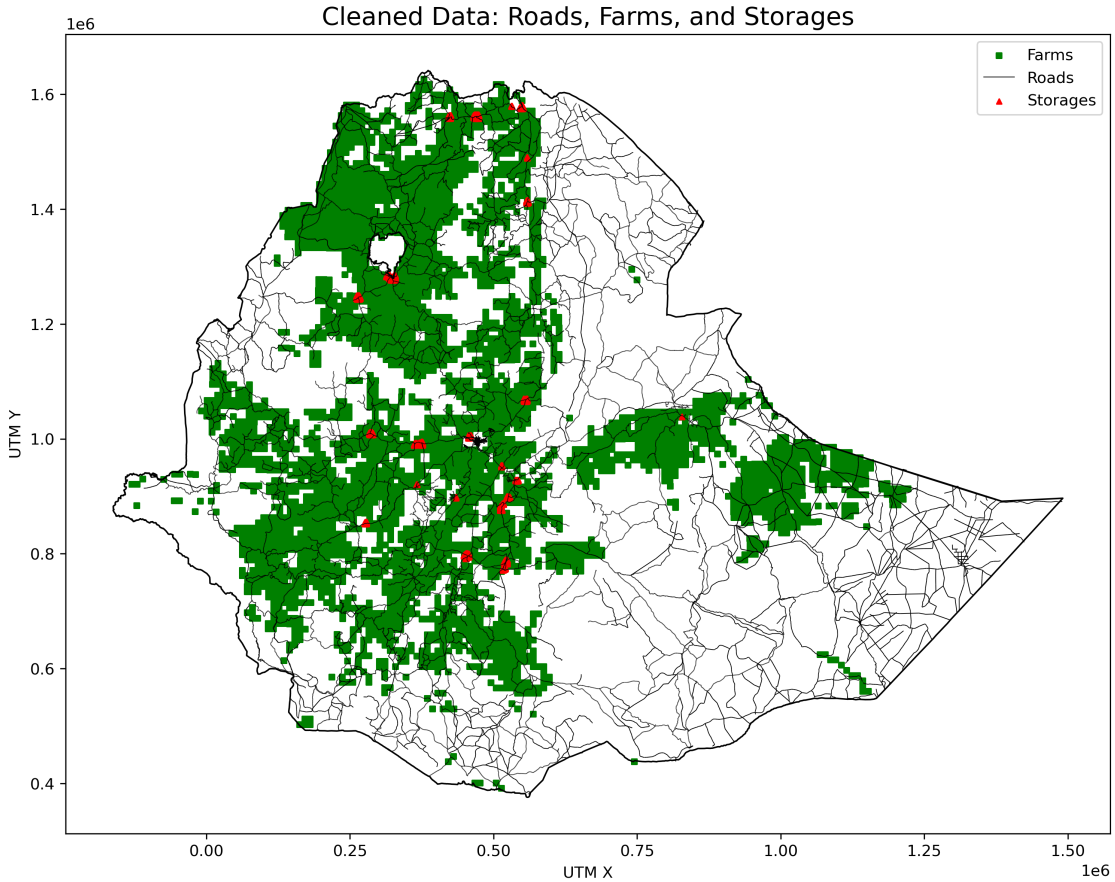


Figure 2.10: Clean Data Map

Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024; Meijer et al., 2018

2.2 Analytical Tools

I conducted this research using Python. Various tools were used for data analysis.

Workflow Management

- Azure was used to create a virtual machine. This enabled me to perform larger computations as my the capacity of my local device had been exceeded.
- venv is Python's built-in module to create virtual environments. I used to create a lightweight environment for my project. This ensured that my packages were isolated, preventing conflicts (Python Software Foundation, 2024).
- Jupyter Notebook is a platform for data analysis as it integrates code with visualiza-

tions, allowing for data exploration to be more integrated (Thomas et al., 2016).

Data Processing and Management

- Pandas efficiently handles 2-dimensional datasets (pandas development team, 2020). The DataFrame data structure easily translates tabular sources. Primarily, this tool was used in the cleaning and filtering stage of analysis.
- NumPy provides efficient array processing for large-scale computations (Harris et al., 2020).
- Fiona was used to simplify GeoJSON files used for transportation maps (Gillies et al., 2024).

Geospatial Analysis

- GeoPandas extends the capabilities of the Pandas package to support geospatial data. This was used to manipulate and join different datasets (Jordahl et al., 2020).
- Rasterio was used to efficiently read raster datasets, like those used in the farm mapping data (Gillies, 2024a).
- Shapely assisted in performing geometric operations (Gillies, 2024b).

Data Visualization

- Matplotlib was used to generate graphs (Hunter, 2007).
- Contextily allowed me to integrate base maps on geospatial plots to understand where different data points are located within Ethiopia (Vincent & Haldane, 2024).

Statistical and Computational

- NetworkX package allowed me to construct and further analyze the network (Hagberg et al., 2008).
- CKDTree is a tool from SciPy, which is designed for spatial queries. It is useful when finding the nearest neighbors and other proximity-based calculations in multi-dimensional space (Virtanen et al., 2020).

Optimization

- PuLP is a Python library for defining linear and mixed-integer programming problems

(Mitchell et al., 2011). It was originally used for optimizing farm-to-storage assignments, relying on the default CBC solver.

- CBC Solver is a open-source solver for mixed-integer programming. It performs well for smaller optimization problems, but it struggles with more complex ones (Forrest et al., 2024).
- Gurobi is a commercial optimization solver. Due to its speed and ability to handle non-convex objective functions, Gurobi was ultimately chosen to handle the more complex and large-scale optimization problems (Gurobi Optimization, LLC, 2024).

2.2.1 Constructing the Network

The network structure used in this analysis is called a bipartite network. A bipartite network specifies that nodes of the same type cannot be connected to each other. We use this structure to model the network of farm and storage nodes because, in this supply chain optimization, we examine the relationships between these two layers. Projections of this graph can then be made, only including nodes of the same type, with links where they share edges with the opposite type in the original graph. Figure 2.11 exhibits this structure.

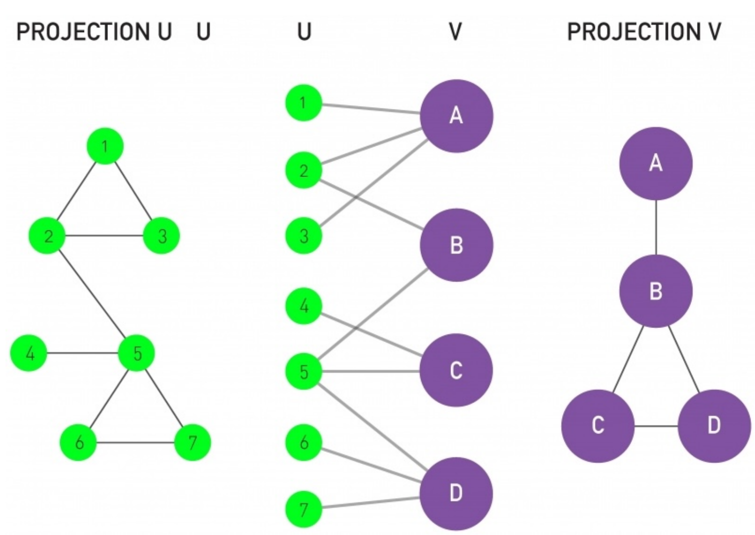


Figure 2.11: Bipartite Network Structure
Image Sourced From: Barabási, 2016

This bipartite network consists of two node types: farm and storage. In Figure 2.12, represents farm nodes in blue, while the crop storage locations are in red. The storage node representation in this graph can be misleading because it looks like there are only a few. However, there are 853 nodes representing potential storage locations. They are mapped on a 1 km grid-cell, which is difficult to visualize on such a large map.

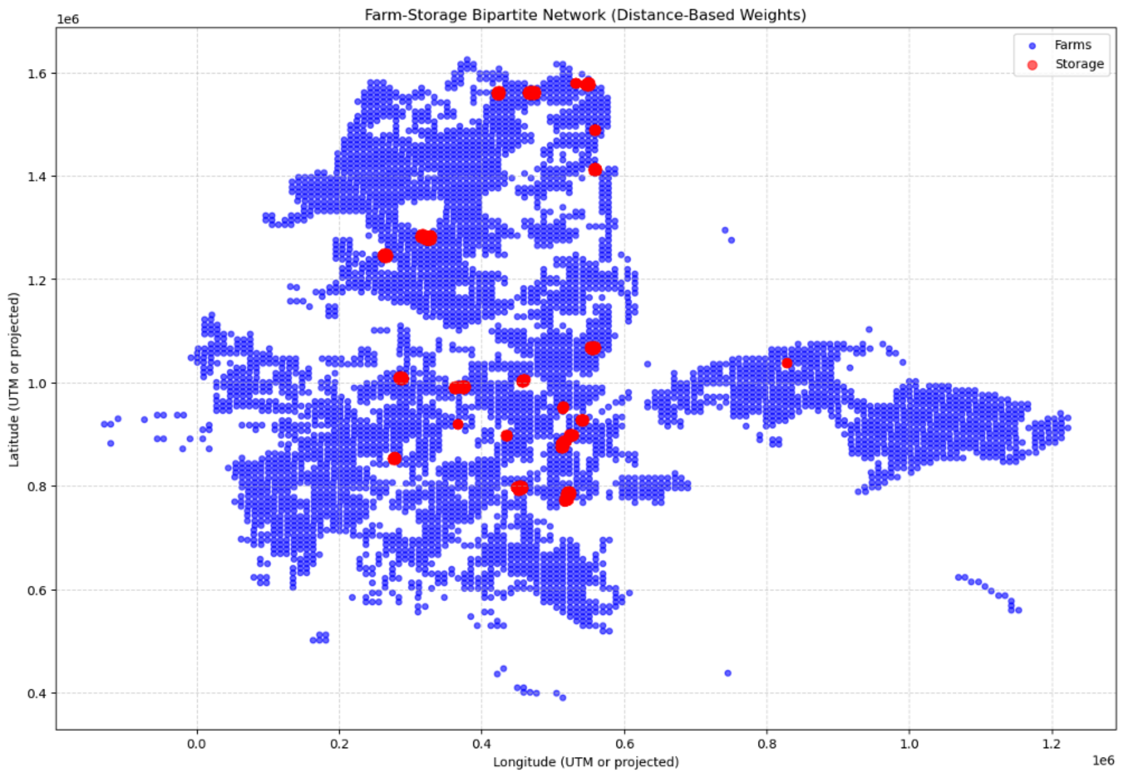


Figure 2.12: This graph represents the nodes of the network.

Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024

In the current model, each farm node is connected to all storage nodes. Since there are 853 potential storage locations identified by the FAO and 4316 farm nodes, this makes the total number of edges (connections) in this graph over 3.5 million. It is necessary to visualize a subset of the dataset because all edges cannot be effectively visualized. We examine a subset of 50 and 1000 farm nodes, each connected to the nearest 10 storage nodes. Figure 2.13 visualizes the subsets of 50 and 1000 farm nodes.

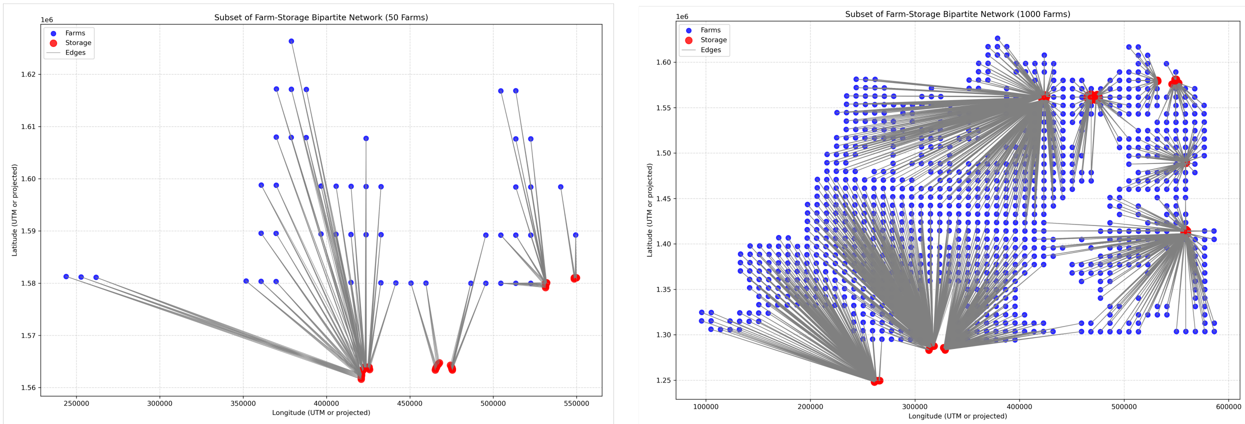


Figure 2.13: Graphed subset of 50 vs 1000 farm nodes.
Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024

2.3 Optimization Problem

The analysis starts with a simplified, base-case model for the storage location optimization problem. Later sections introduce additional complexity by relaxing various constraints for better approximations to reality. First, the model only optimizes locations for maize, thus abstracting away from different storage needs and costs of different cereals.

Storage facilities exhibit high construction costs, thus the model should optimize the locations for a realistic number of facilities. The FAO suggests 25 distinct, concentrated contiguous regional areas for potential storage facility locations. Thus, as a natural simplification, the model will consider these concentrated areas as one potential storage location, implicitly restricting the number of potential storage facility nodes to 25. This is a modest restriction since, for practical purposes, one could think of 20 storage facilities located in close proximity to one another as a single storage location with 20 times the capacity of a representative storage facility.

The objective of this optimization model is to identify final cereal crop storage facility locations, which minimize the total transportation cost from farm areas to storage facilities.

The social planner's objective is to choose storage locations that minimize aggregate transportation costs as a function of the distances between farm nodes and their assigned storage storage location. Equation (2.1) specifies the objective function.

$$\min \sum_{f \in F} \sum_{s \in S} x_{f,s} \cdot d_{f,s} \tag{2.1}$$

The variables, parameters, and constraints are as follows:

Variable and Parameter Notations

$$\begin{aligned}
F &: \text{Set of 4043 farm nodes producing maize (indexed by } f \in F) \\
S &: \text{Set of potential storage nodes (indexed by } s \in S) \\
p_f &: \ \text{Annual production of maize at farm } f \\
(a_f, b_f) &: \ \text{Coordinates of farm } f \\
(a_s, b_s) &: \ \text{Coordinates of storage node } s \\
d_{f,s} &: \ \text{Euclidean distance between farm } f \text{ and storage node } s, \\
&\quad \text{calculated as } d_{f,s} = \sqrt{(a_f - a_s)^2 + (b_f - b_s)^2} \\
M &: \ \text{Maximum number of facilities that can be built}
\end{aligned}$$

Decision Variables

$$x_{f,s} = \begin{cases} 1 & \text{if farm } f \text{ sends its entire maize production to storage } s, \\ 0 & \text{otherwise.} \end{cases}$$

$$y_s = \begin{cases} 1 & \text{if facility } s \text{ is built,} \\ 0 & \text{otherwise.} \end{cases}$$

Constraints

1. *Assignment Constraint:* Each farm must send its maize to exactly one storage facility.

$$\sum_{s \in S} x_{f,s} = 1 \quad \forall\, f \in F \tag{2.2}$$

2. *Facility Existence Constraint:* Farms can only send to a storage location if it is built.

$$x_{f,s} \leq y_s \quad \forall\, f \in F,\ s \in S \tag{2.3}$$

3. *Facility Limit Constraint:* At most M facilities can be constructed.

$$\sum_{s \in S} y_s \leq M \tag{2.4}$$

4. *Market Clearing Constraint:* All maize production must be stored at some facility.

$$\sum_{f \in F} \sum_{s \in S} x_{f,s} \cdot p_f = \sum_{f \in F} p_f \tag{2.5}$$

Assumptions and Limitations

The proposed system imposes several assumptions. First, each of the farm nodes transports output directly to the final crop storage facility. The model restricts the number of local storage and distribution centers because of data availability. While these assumptions facilitate further analysis, they impose limitations on the degree to which the model approximates realistic settings. The following list summarizes the limitations corresponding the structure of the model:

- Assignment Constraint: This constraint forces each farm to send their crop to exactly one storage facility. This will result in remote farms transporting cereals for very long distances, even if they have a small amount of production.

- Market Clearing Constraint: Currently, it is assumed that all cereal farms are connected to the market. It is not realistic for all production to end up at a final crop storage location. These storage locations are only meant for product entering the market. It is estimated that 20 to 30 percent of maize makes it to the market in Ethiopia (Rashid et al., 2010). One 2022 study of maize farmers in southwest Ethiopia found that out of 546 maize farmers, 381 were market participants while 165 were not participants (Haile et al., 2022b).

- Euclidean Distance as Cost: This model assumes transportation cost per ton between each farm node and storage facility will be equal to the straight-line distance. This ignores the current transportation infrastructure of Ethiopia.
- Storage Facility Capacity: This benchmark does not enforce capacity limits for storage facilities. This could cause unrealistically large amounts of production to be assigned to some storage nodes.

One key issue is the lack of agricultural data in Sub-Saharan Africa. A paper on reducing postharvest losses in Ethiopian cereals urges that improving data collection is crucial for identifying and implementing targeted solutions to reduce postharvest losses (Hengsdijk & de Boer, 2017, p. 945). The yield data used for this model is based on an average between the years 2019 and 2020. It does not specify the variation in yield between seasons. This is an important factor to consider when looking at optimizing storage facilities, which have capacity maximums. For example, during the off-season, it would be helpful to know if some of these storage facilities should stop operating.

One major limitation of this research is the lack of farm entity-specific data. There is information on how different crop yields are geo-spatially distributed. However, there is data at the farm-level. It is an assumption that these different grid-cells would act as individual nodes on this network.

Additionally, there is not information on where local, smaller storage and distribution centers exist. There isn't a full mapping of where the current final crop storage and distribution centers exist. This prevents an accurate representation when constructing a cost function. Some facilities would simply need to be renovated with more modern storage technology while others would need to be constructed from the ground up.

The lack of supply chain data prevents an extension of this network past the suggested final ceral crop storage locations. There is not data on where these goods are further transported. Because Ethiopia is a net importer of agricultural goods, there isn't much information that can be gathered from sources like the Food and Agriculture Organization of the United Nations Detailed Trade Matrix, which specifies international trade information for agricultural commodities. Additionally, Ethiopia is land-locked, so it does not have direct access to major ports, which would provide additional information and could be used to add another layer to the supply chain network.

Chapter 3: Model Estimation and Results

3.1 Baseline Model

FAO Suggested Locations: These storage locations are calculated based on a subset of locations recommended by the Food and Agriculture Organization (FAO) of the United Nations. They are selected for their access to resources like financing and internet. While this is helpful at looking at the current state of the Ethiopian cereal crop industry, it is important to understand how these storage locations would change if these infrastructure limitations did not exist. This would be the case where financing and access to communications are widespread through Ethiopia. This base case would also be able to give information on the cost of this lack of infrastructure. Doing this can allow us to better understand how these frictions are impacting the overall efficiency of the supply chain. This can be done by looking at all gridcells within Ethiopia instead of just the suggested ones and performing the same linear optimization. The total "cost" of transportation can be compared to that of the system using the suggested FAO locations.

50x50 km Gridcells: In this model, there is one farm node for each 50x50 km grid cell of land in Ethiopia that produces maize. Each storage node represents one 50x50 km grid cell of land within Ethiopia. This means that everywhere in Ethiopia is being considered as a potential storage location. Disregarding the FAO suggested locations alters the optimization problem considerably.

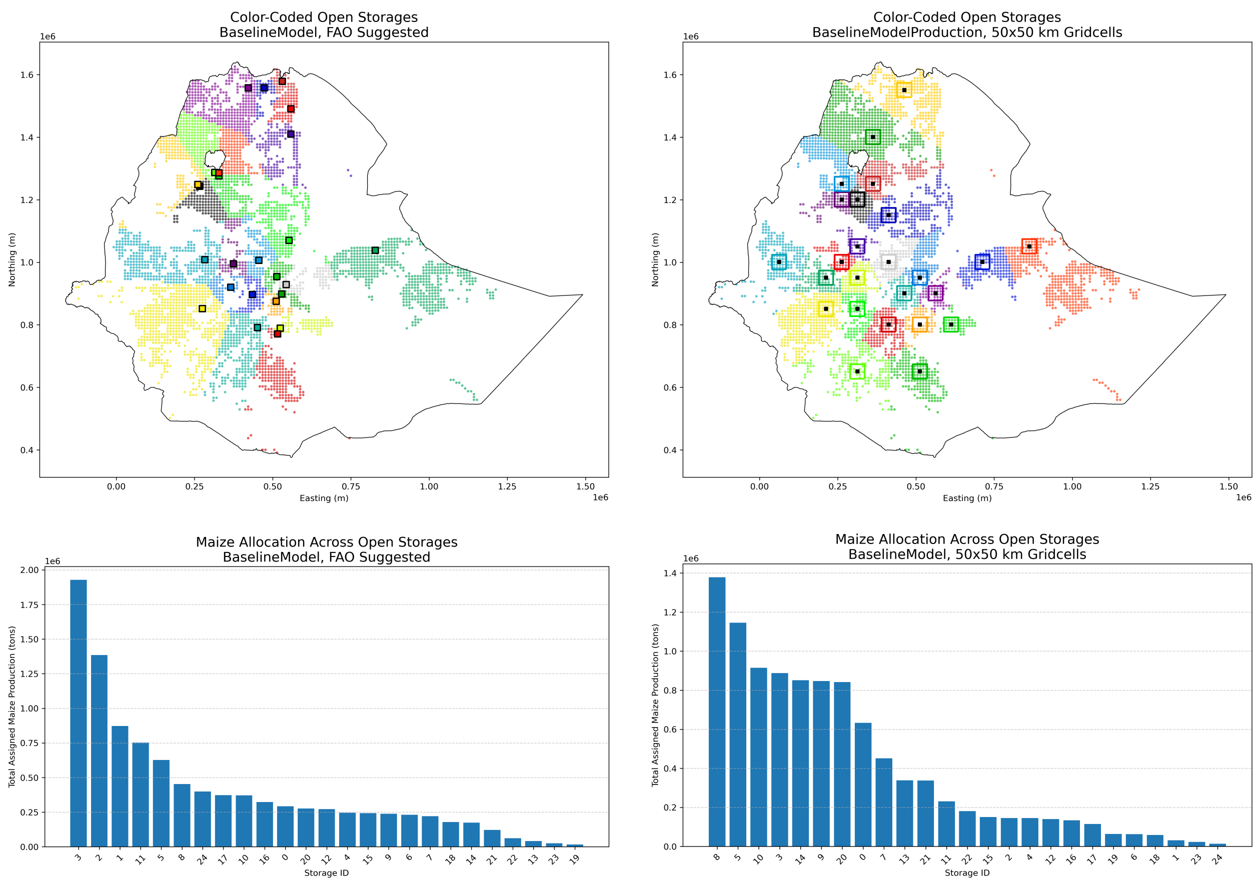


Figure 3.1: Baseline Model Results

Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024

3.2 Production-Weighted Euclidean Distance

The farm nodes have varying levels of production. It is inaccurate to assume that a farm producing far more maize than another would have equal transport cost for the same Euclidean distance. Equation (3.1) specifies the objective function, where the Euclidean distance is now production-weighted.

$$\min \sum_{f \in F} \sum_{s \in S} x_{f,s} \cdot d_{f,s} \cdot p_{f,s} \tag{3.1}$$

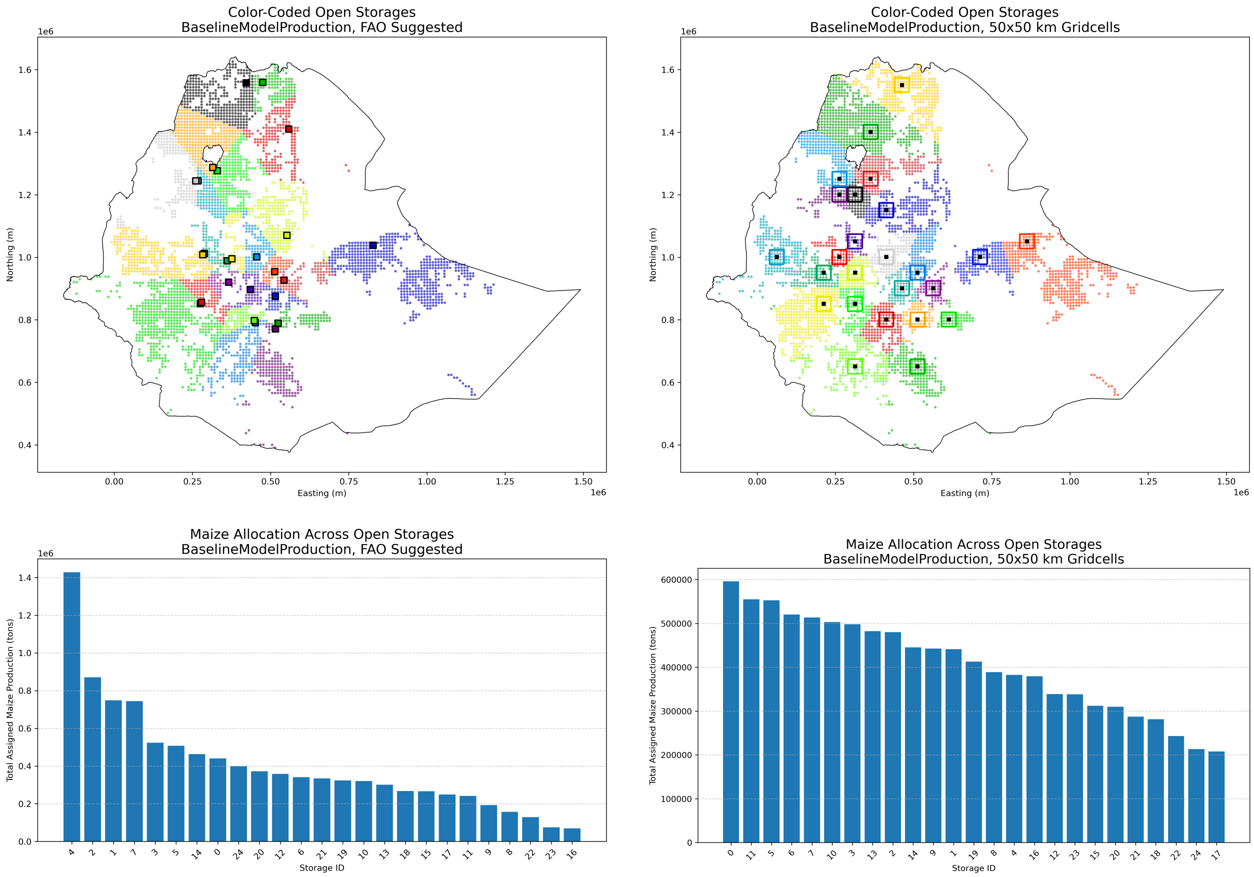


Figure 3.2: Production-Weighted Baseline Model Results

Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024

3.3 Travel Time Model

The prior models specify transportation cost as a function of the Euclidean distance between storage facilities. However, this specification oversimplifies physical limitations of the transportation infrastructure in Ethiopia. Effectively, minimizing the Euclidean distance between nodes assumes that there is a straight rode that directly connects each farm to the potential set of matching facilities.

To better approximate reality, we restrict each farm and storage node to travel via the existing road infrastructure in Ethiopia. Roads are a continuum of nonlinear latitude and longitude coordinates, typically constrained by physical limitations such as rivers and mountains. Thus, we assume that a farm's starting travel point begins at the nearest road coordinates in terms of Euclidean distance. This assumption better approximates reality, subject to the limitations of the data for private local roads. We assign a travel speed of 40 km/hr to the distance of the nearest road access, which matches the standard speeds in Ethiopia for local and urban roads.

Based on the revised travel constraints, we can further simplify the optimization problem by grouping farm and storage nodes that access the same initial road location. That is, after accessing the road, their optimal travel times will follow the same route. However, their total travel times can still differ by the distance it takes them to travel to their initial road access point. Effectively, once we know where a farm or storage facility access the road infastructure the travel time to get to the road can be ignored for edge optimization purposes, and the centroid of the cluster of nodes with the same access point can be used instead.

Some farm or storage nodes may snap to the same road node. This simplifies calculations because the equivalent node pairs can be treated as identical for the shortest road route calculation. Figure 3.3 shows what this would look like. This is an example of a cluster of storage nodes suggested by the Food and Agriculture Organization of the United Nations. Graphical representations of all snapped storage clusters can be found in the Appendix.

It is computationally expensive to calculate all exact farm-storage distance pairs. This is particularly problematic with the set of FAO Suggested storage nodes. In order to simplify the shortest path calculation, the FAO Suggested storage nodes are grouped together by their regional cluster. I defined the centroid of each FAO Suggested storage cluster as the average of all (x, y) values of the individual storage nodes. The centroid can then be snapped to the nearest road node.

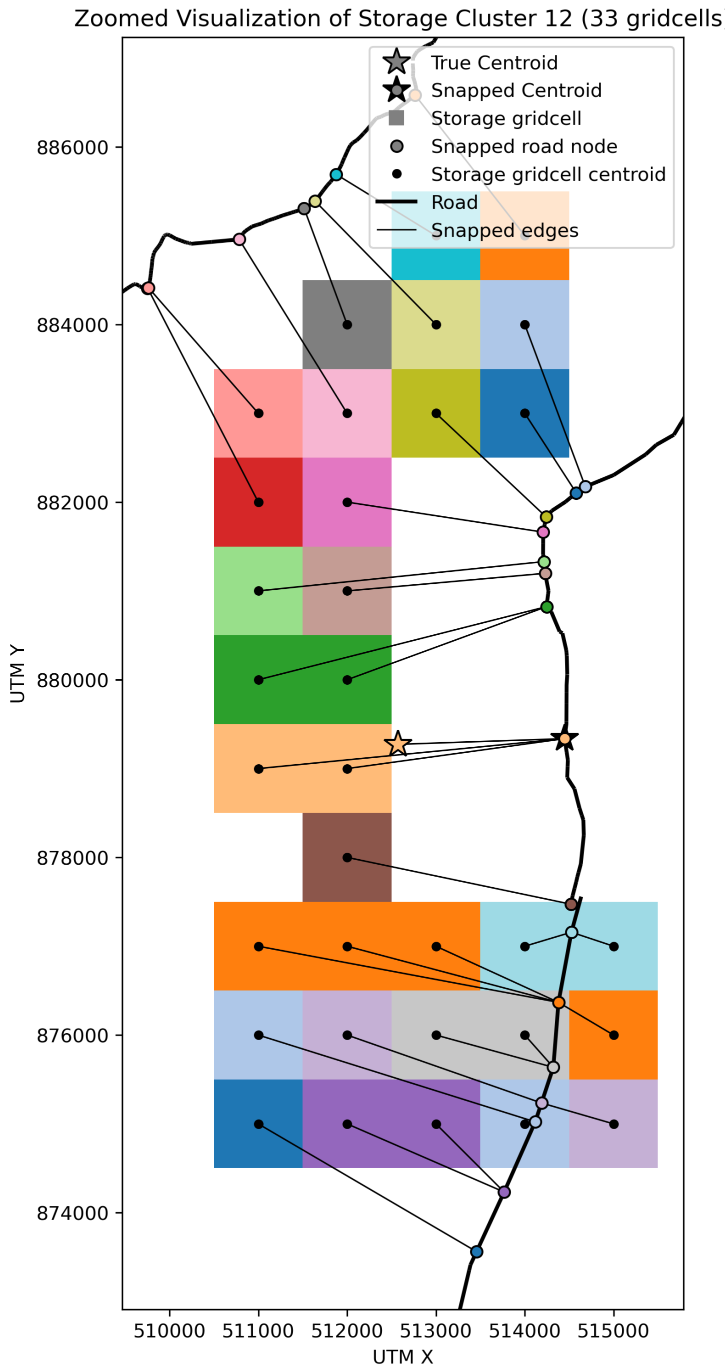


Figure 3.3: Storage snapping to road nodes.
Data Source: Food and Agriculture Organization of the United Nations, 2022; Meijer et al., 2018

Figure 3.4 displays a validation step between this shortest path calculation and Google Maps. The model performed very well in identifying the shortest road path between random storage and farm nodes.

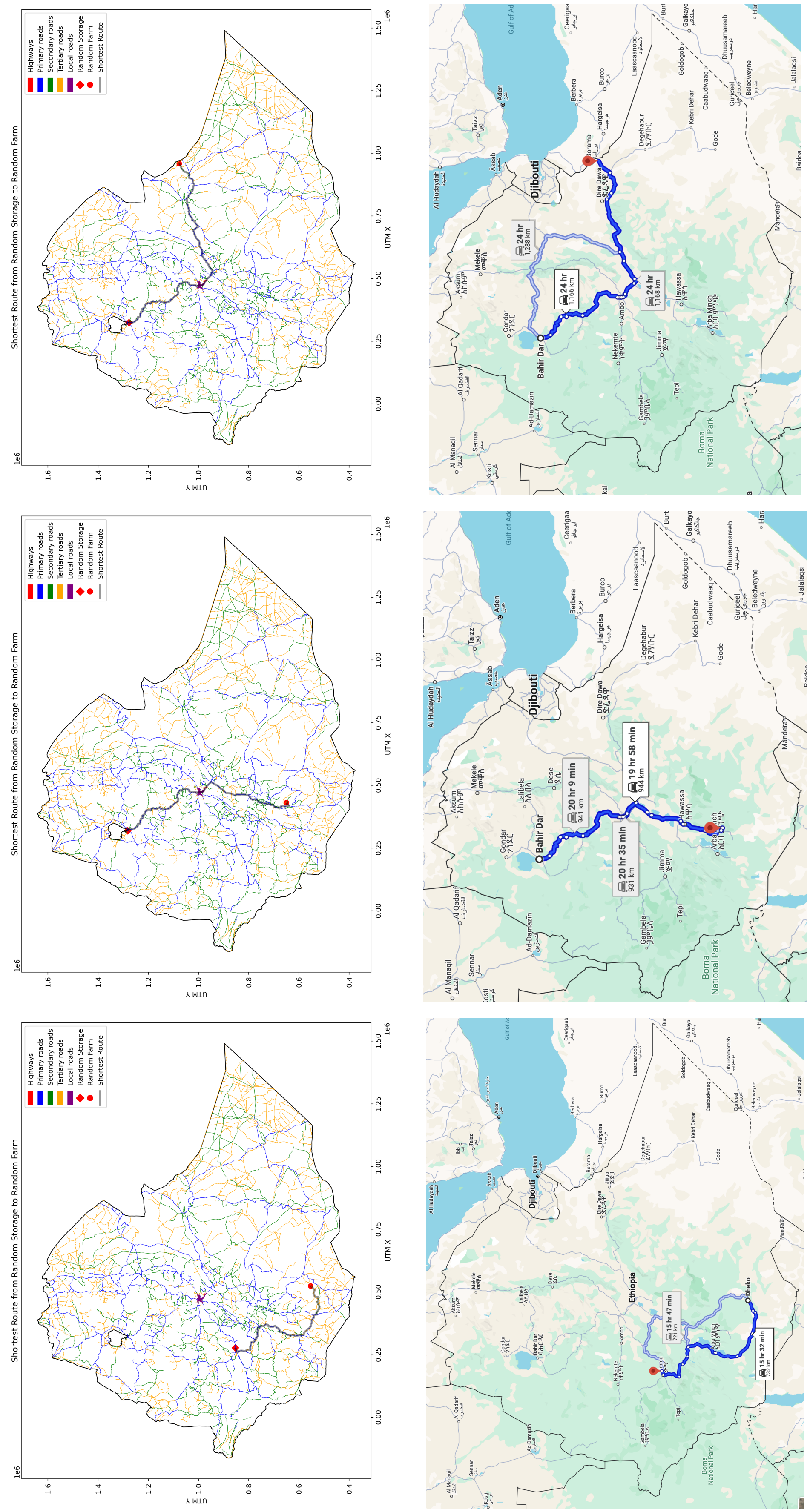


Figure 3.4: Transport Cost Validation with Google Maps

Data Source: (Food and Agriculture Organization of the United Nations, 2022; Google, 2025; International Food Policy Research Institute (IFPRI), 2024; Meijer et al., 2018)

After finding the shortest path, the segment lengths are multiplied by varying speeds for each road type. These estimated speeds are: 100 km/hr for highways and primary roads, 60 km/hr for secondary roads, 50 km/hr for tertiary roads, and 40 km/hr for local and urban roads (HPS, 2025; WorldData.info, 2025). Figure 3.5 maps the nearest farm nodes to a chosen storage node by travel time. This displays the difference between the travel time calculation and simple Euclidean distance.

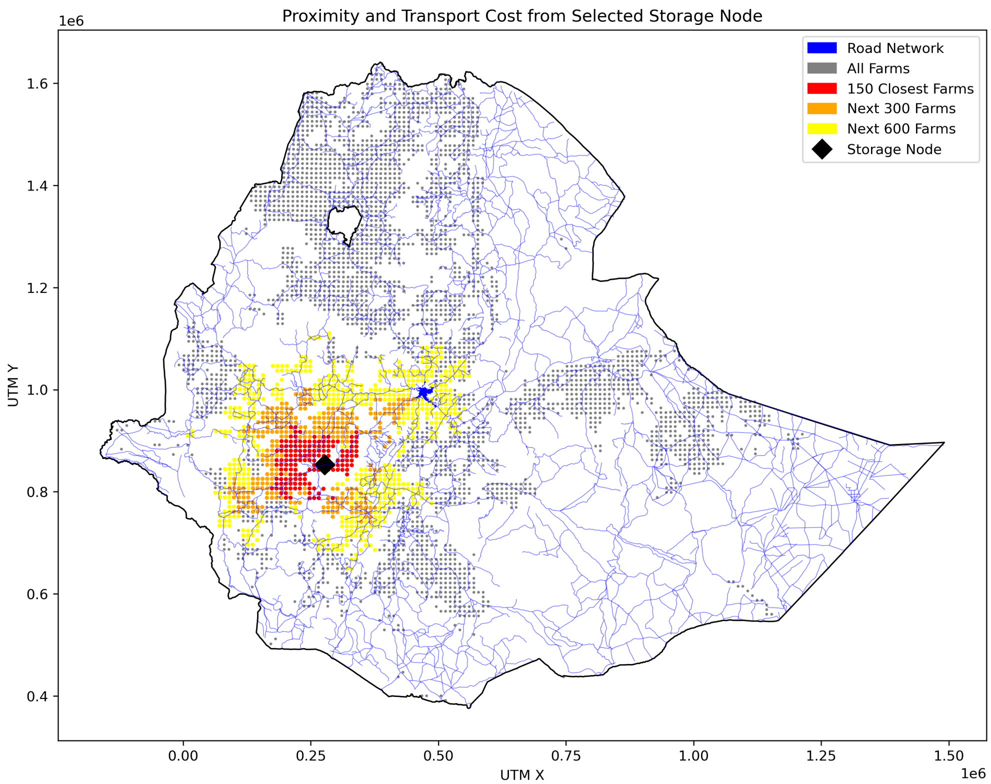


Figure 3.5: Closest farm nodes to selected storage node.
Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024; Meijer et al., 2018

For this model, we minimize calculated travel time instead of Euclidean distance. The objective function is specified in Equation (3.2)

$$\min \sum_{f \in F} \sum_{s \in S} x_{f,s} \cdot t_f \cdot p_{f,s} \tag{3.2}$$

$t_{f,s}$: calculated travel time from farm node f to storage node s

Figure 3.6 visualizes the production-weighted travel time model results.

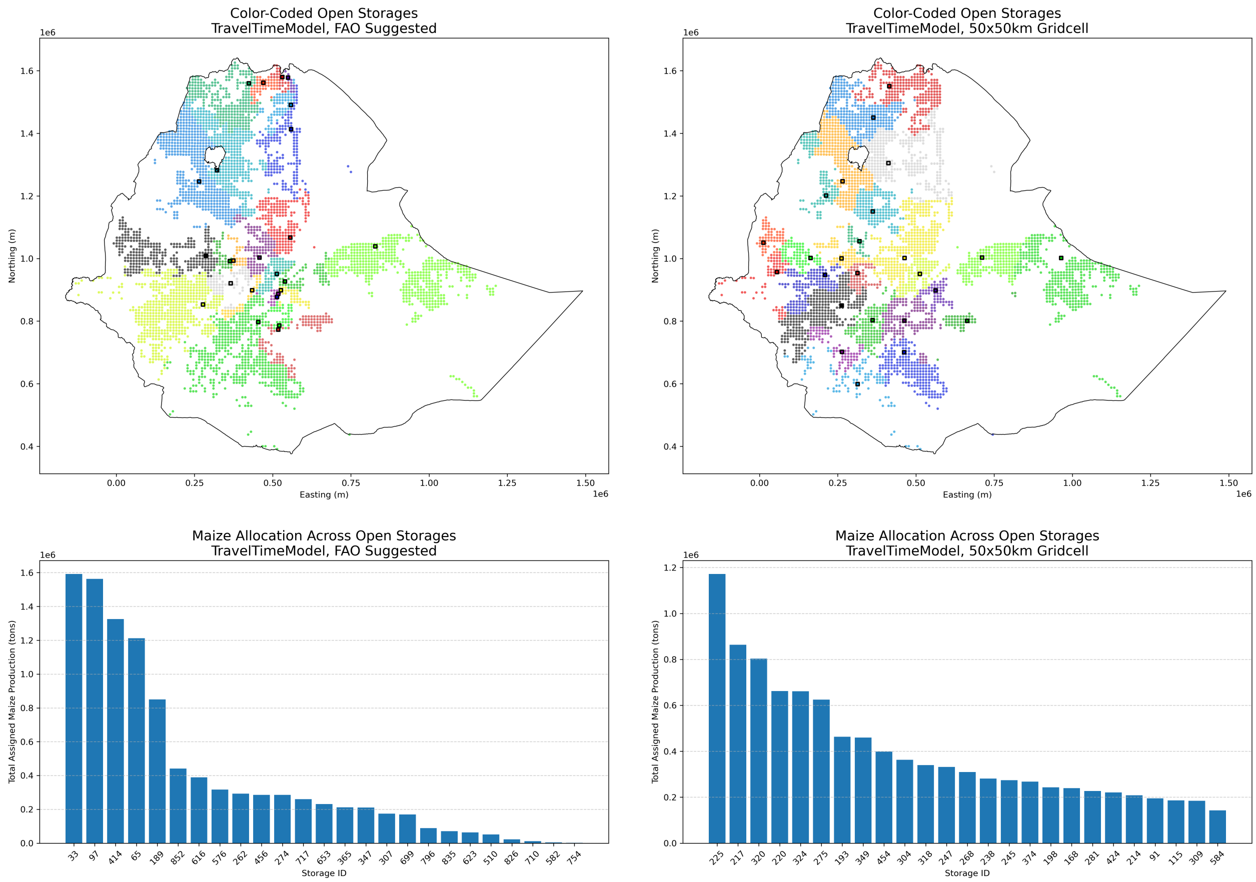


Figure 3.6: Production-Weighted Travel Time Model Results

Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024; Meijer et al., 2018

3.4 Load Balancing Model

The previous models do not limit yield being sent to each storage facility. For the FAO Suggested Storage version of the Baseline Model, there are nearly 2 million tons of crop being sent to one facility, which is unreasonable. However, we can build on the previous model to incorporate a load balancing term, penalizing uneven distribution of maize across storage facilities. This helps ensure that no single storage is overloaded.

The *Load Variance Definition Constraint* is created:

$$V_s = \left| \sum_{f \in F} p_f \cdot x_{f,s} - \frac{\sum_{f \in F} p_f}{M} \right|, \quad \forall\, s \in S \tag{3.3}$$

The first term is the total production going to storage s. The second term is the desired production load for each storage, if they production were to be evenly distributed across all storages. The second term is the total production for all farms divided by the number of storage facilities.

The objective function is adjusted to:

$$\begin{gathered} \min \sum_{f \in F} \sum_{s \in S} x_{f,s} \cdot t_f \cdot p_{f,s} \;+\; \lambda \sum_{s \in S} V_s \\ \lambda :\ \text{load balancing penalty term} \end{gathered} \tag{3.4}$$

This objective retains the structure from the previous model, which aims to minimize production-weighted travel time. However, a penalty term for load variance is added. λ is the weight of the load balancing penalty term. In the model results represented in Figure 3.7, λ is set to 0.97.

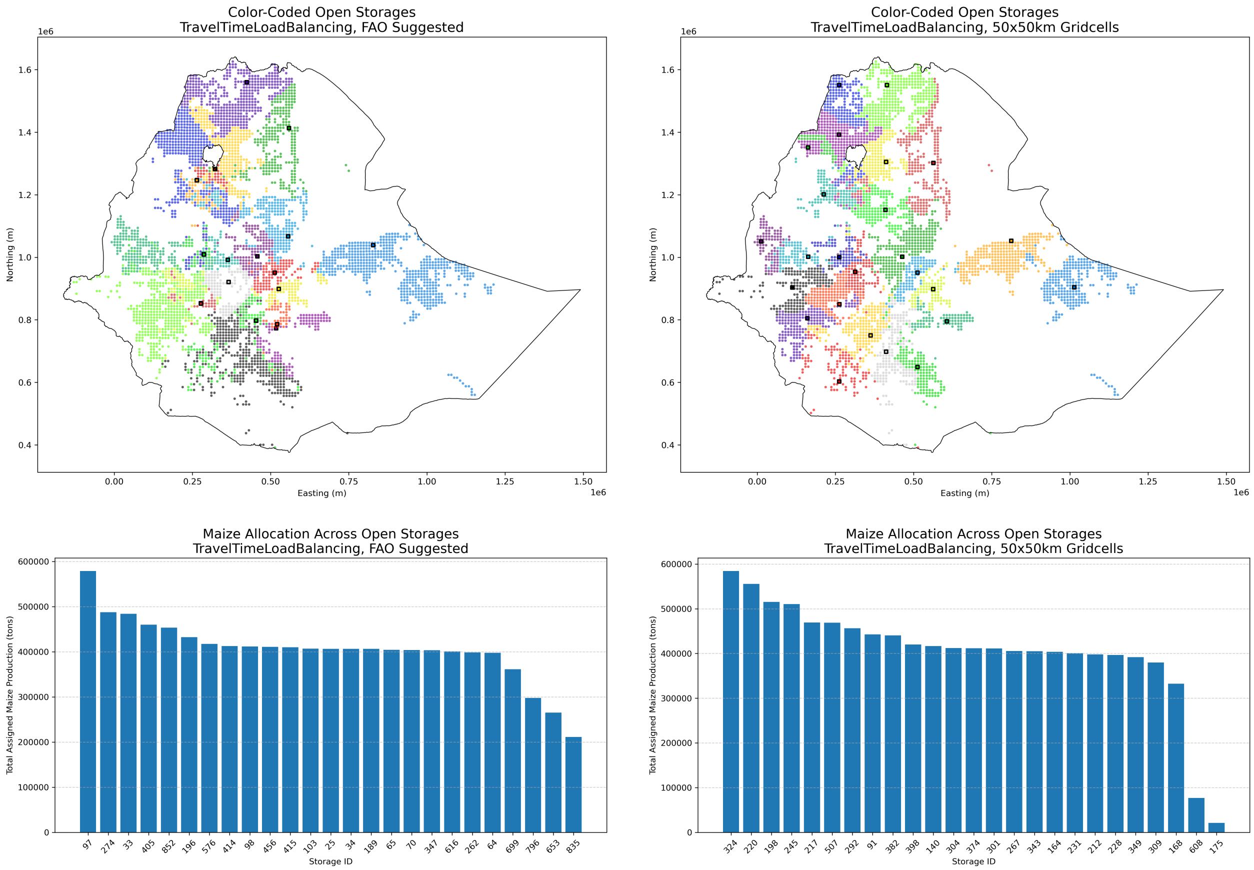


Figure 3.7: Linear Load Balancing Model Results

Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024; Meijer et al., 2018

Load balancing can be done linearly or quadratically. A linear load-balancing model, as shown above, penalizes any deviation from the desired load equally. A quadratic model penalizes greater deviations more. Linear optimizations are much simpler for optimization programs to solve. However, the production distribution achieved with a quadratic load-balancing model is desirable. For a quadratic load-balancing optimization, the objective function can be adjusted to:

$$\min \sum_{f \in F} \sum_{s \in S} x_{f,s} \cdot d_{f,s} + \lambda \cdot \left[\sum_{s \in S} \sum_{f \in F} p_f \cdot x_{f,s} \right]^2, \quad \forall s \in S \tag{3.5}$$

The first term of the objective function minimizes total Euclidean distance, while the second term penalizes unbalanced load across storage locations. The quadratic penalty term

L_s^2 discourages disproportionately large production allocations, where L_s is the amount of production allocated to each storage facility. λ is the weight of the quadratic load balancing penalty term. A sensitivity analysis can be run to identify a λ suiting our preferences for load balancing.

This is a quadratic objective due to the L_s^2 term. Because of this increased level of complexity, the computational power required to run this optimization is greatly increased from the baseline. So far, I have been able to run this for the set of potential storage locations consisting of all 150x150 km gridcells of Ethiopia, with a λ value of 0.0001. While this is not a good representation of what the storage locations should be, since the set is so small, it is a good representation of what the distribution of production should look like.

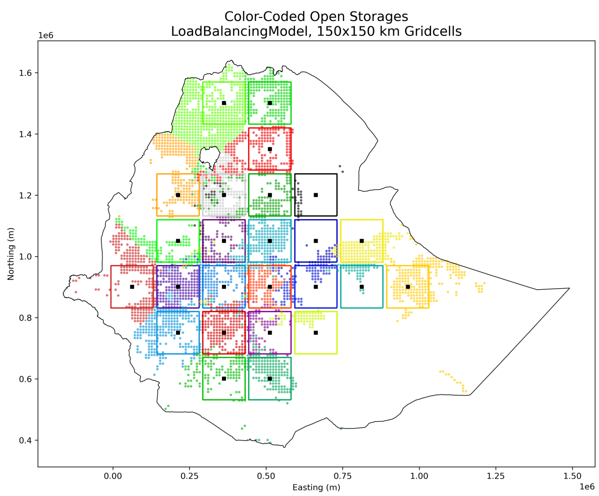


Figure 3.8: Color-Coded Assignment

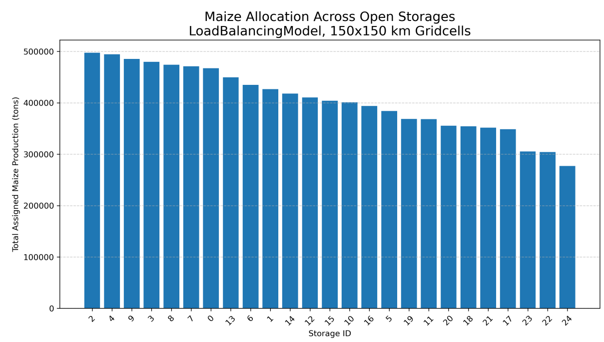


Figure 3.9: Production Distribution

Figure 3.10: Quadratic Load Balancing Model Results

Data Source: International Food Policy Research Institute (IFPRI), 2024

3.5 Preliminary Analysis of A Resilience Model

This section presents a discussion on preliminary analysis of a resilience model. While presenting formal results would be premature, an informal discussion may facilitate additional insights from the prior analysis. Future work will include a more complete development and estimation of the resilience model.

When farms are only reasonable serviced by one storage facility, it is not a resilient structure. If that facility becomes compromised, it would be a high-cost solution to send crops from those farms elsewhere. Percolation analysis can help address this issue. Percolation theory can be used to test for network robustness. This technique removes one node from a network and observes the damage. (Barabási, 2016)

Outside of an optimization setting, there are two ways to do this: conduct random node removal (simulating failures) and remove important hubs (simulating attacks). In this case, both could be important. Unexpected climate events could cause failures in the network. Worker strikes could target important storage locations.

In order to add resilience to the objective function of the baseline model, we can calculate resilience cost. Resilience cost can be calculated differently whether you are optimizing for attack or failure resilience. Let $d'_{f,s}$ represent the reroute distance for farm f if storage s fails. This represents the distance to the nearest alternative storage. The new objective function for this model would be to minimize the weighted sum of transportation and resilience costs, with α and β acting as respective weights.

Attack Resilient Model: To ensure that the solution is robust against targeted attacks, we require that the resilience cost, z_{attack}, is at least the aggregated reroute cost for every storage facility. This minimizes the worst-case (maximum) reroute cost for any storage failure.

$$z_{\text{attack}} \geq \sum_{f \in F} x_{f,s}\, d'_{f,s} \quad \forall\, s \in S. \tag{3.6}$$

The corresponding objective function is

$$\min\ \alpha \sum_{f \in F} \sum_{s \in S} x_{f,s}\, p_f\, d_{f,s}\ +\ \beta\, z_{\text{attack}}, \tag{3.7}$$

where $\alpha = 0.8$ and $\beta = 0.2$ (for example) prioritize transport cost over resilience.

Failure Resilient Model: Alternatively, to simulate random failures, we define the failure resilience cost as the average aggregated reroute cost across all open storage facilities. In this case, all storage facilities are equally likely to fail.

$$z_{\text{fail}} = \frac{1}{|S|} \sum_{s \in S} \sum_{f \in F} x_{f,s} \, d'_{f,s}. \tag{3.8}$$

Then the objective function becomes

$$\min \ \alpha \sum_{f \in F} \sum_{s \in S} x_{f,s} \, p_f \, d_{f,s} \ + \ \beta \, z_{\text{fail}}. \tag{3.9}$$

The complexity of this model does not allow for simple formulation into an optimization model. It will be a future development. Ultimately, the goal is to develop a model that not only minimizes the usual production-weighted travel time but also incorporates resilience against disruptions.

Chapter 4: Conclusions

The analysis of the model results reveal distinct trade-offs in transportation efficiency and system realism across both the FAO Suggested and Gridcell scenarios. No single model dominates all performance metrics. Instead, each model excels in the metric that it explicitly optimizes. Figure 4.1 compares the various models on different metrics.

The comparison between the FAO Suggested storage locations and the 50$\times$50km gridcell storage locations shows that the gridcell configuration consistently outperforms the suggested locations in achieving its objective. The total cost for FAO suggested storage Baseline Model was 443,287,161.42 meters. Relaxing the 25 facility restriction to 50 x 50 km gridcells, the total cost becomes 120,744,581.13 meters. This is a significant cost difference of 322,542,580.29 meters. Limiting the storage locations to those suggested by the FAO caused the Euclidean cost for the system to increase by 267 percent.

The model results exhibit the poor performance under models minimizing Euclidean distance. When the FAO storage locations were optimized by production-weighted travel time instead of production-weighted Euclidean distance, the travel time for the system was reduced by 1361 hours.

There are many ways we can define optimality. In order to model the real world more accurately, our objective function must be further developed to capture real-world complexities. Refining the objective function to encompass both economic efficiency and system resilience will enhance its practical applicability.

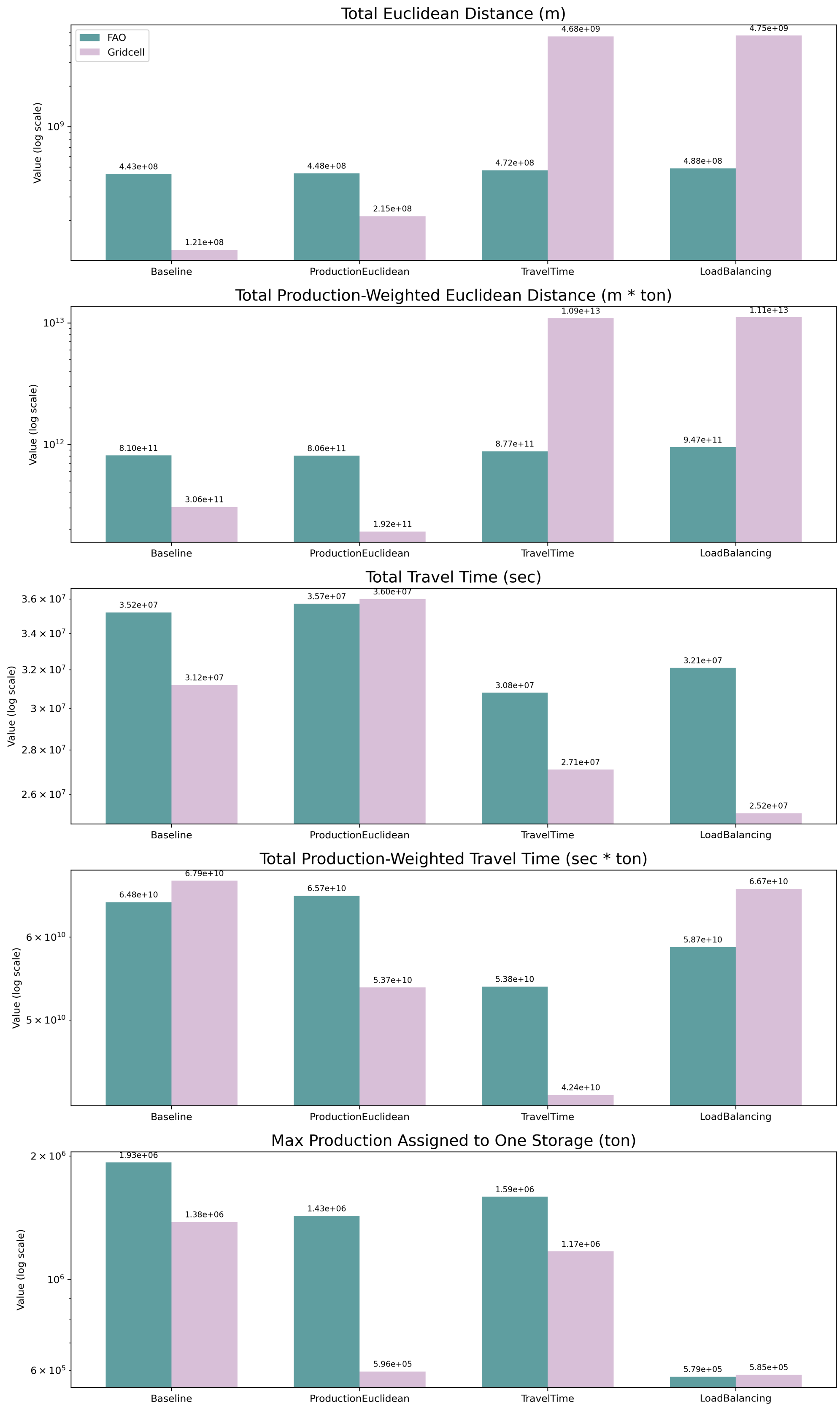


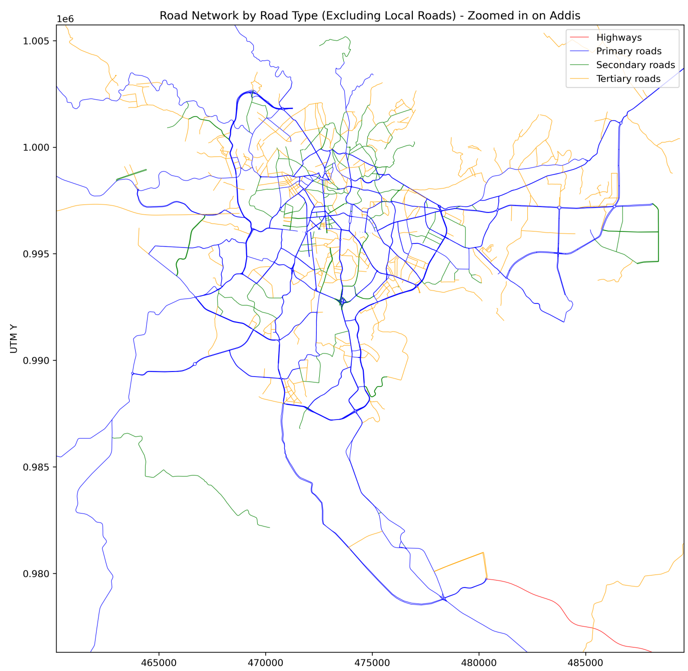
Figure 4.1: Comparison of Models

Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024; Meijer et al., 2018

Appendix

GRIP Road Network Cleaning

Many of the "Local roads" in the urban area of Addis Ababa were disconnected from eachother, causing problems. I verified there were other road types that would allow transport through the region. Then, I filtered to remove this road type.

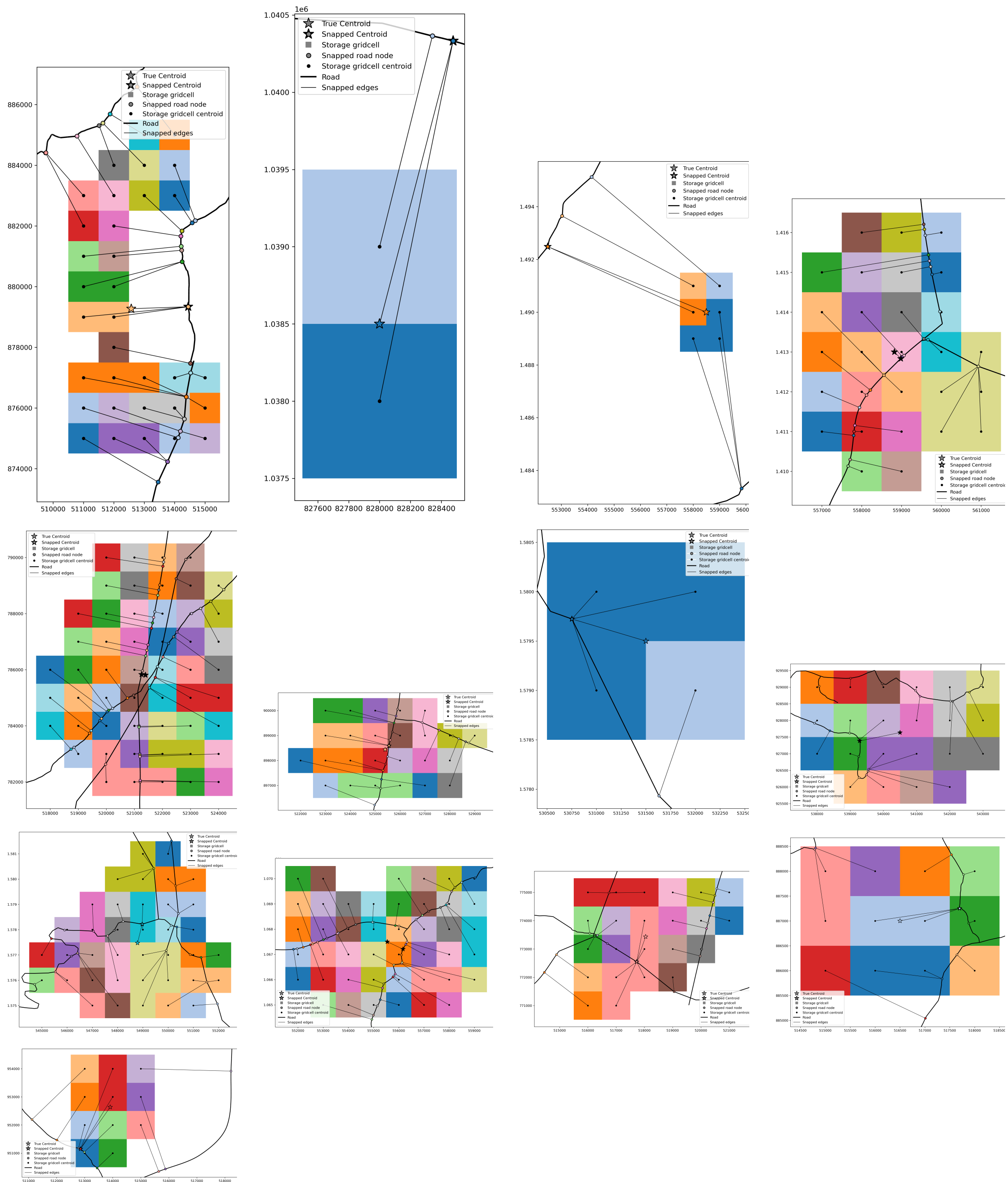


Addis Ababa Road Types (Local Type Removed)

Data Source: Ethiopian Space Science and Geospatial Institute (SSGI), 2020; Meijer et al., 2018

Snapped Storage Clusters

Snapped Storage Clusters (Part 1).



Snapped Storage Clusters (Part 2).

Production Map

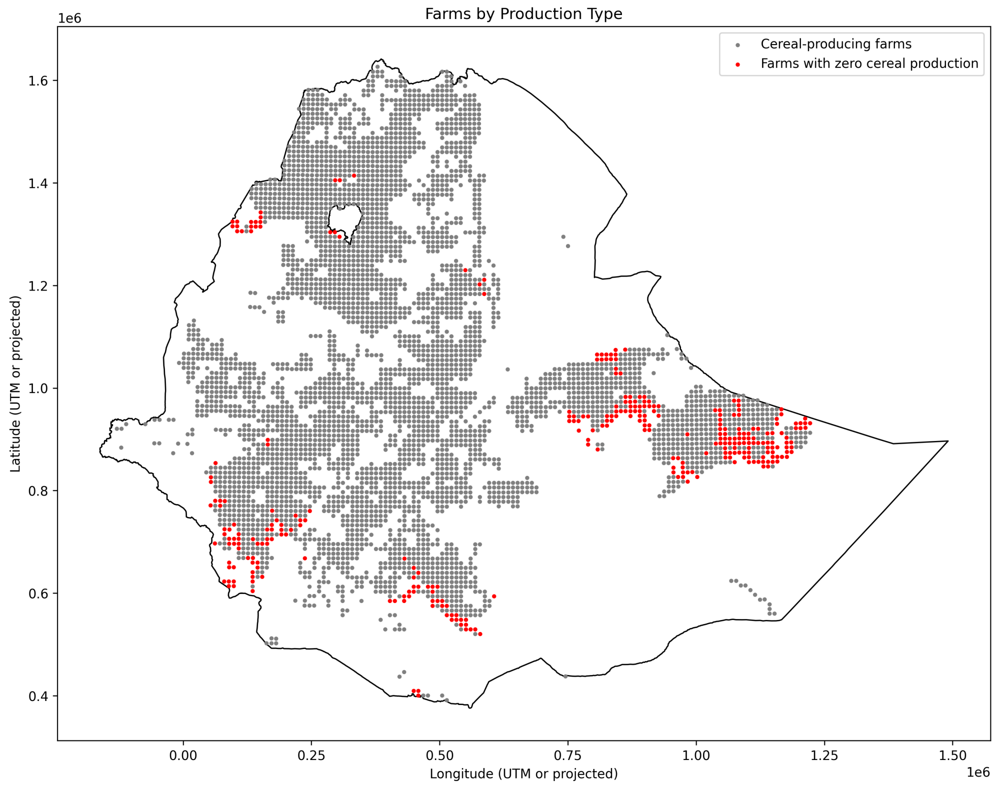


Figure A1: Cereal Producers vs Non Cereal Producers
Data Source: International Food Policy Research Institute (IFPRI), 2024

Cereal Production and Dominance Individually Mapped

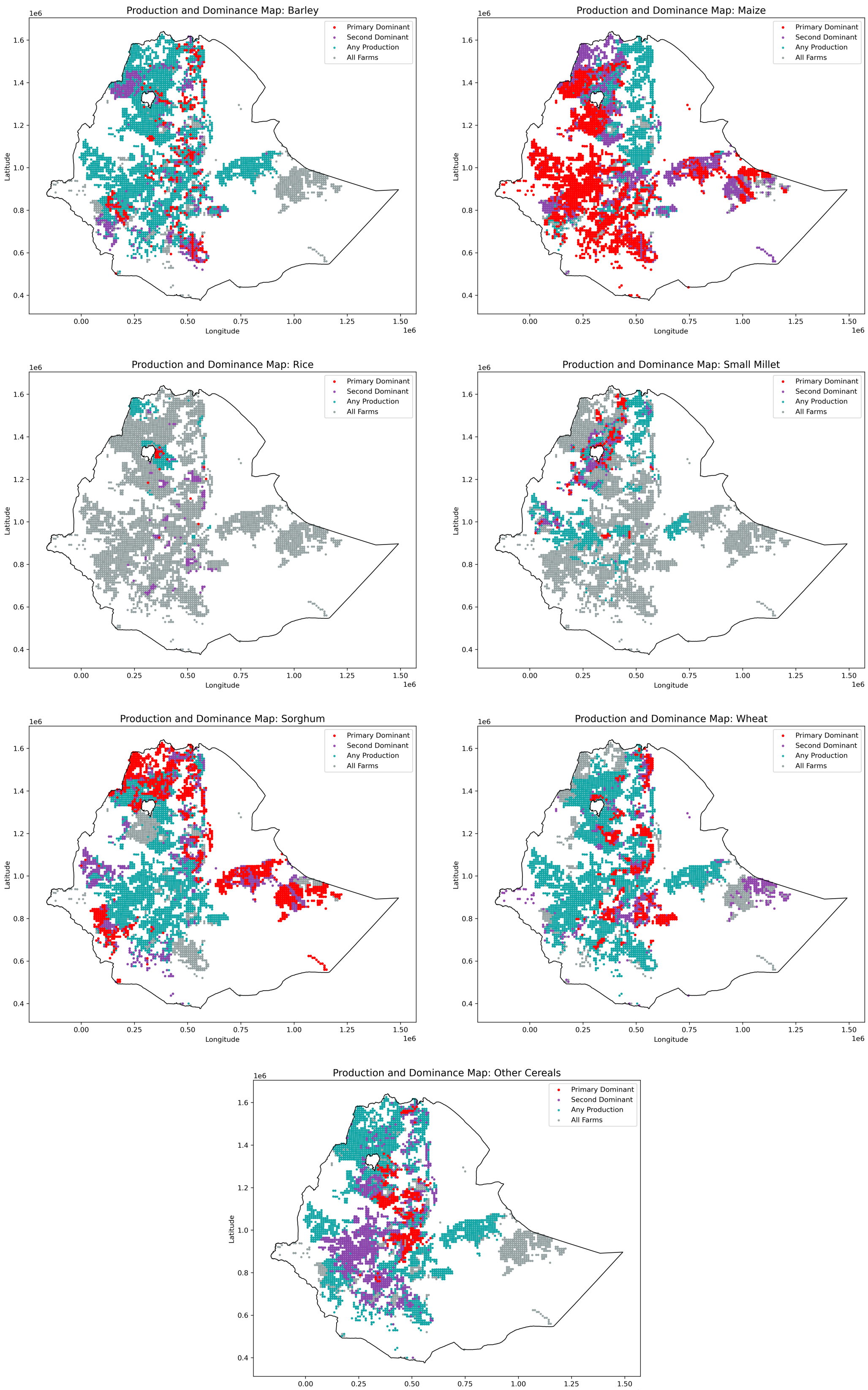


Figure A2: Dominance and production maps for various cereal crops.

Data Source: Food and Agriculture Organization of the United Nations, 2022; International Food Policy Research Institute (IFPRI), 2024; Meijer et al., 2018

Regional Heatmaps of Production by Crop

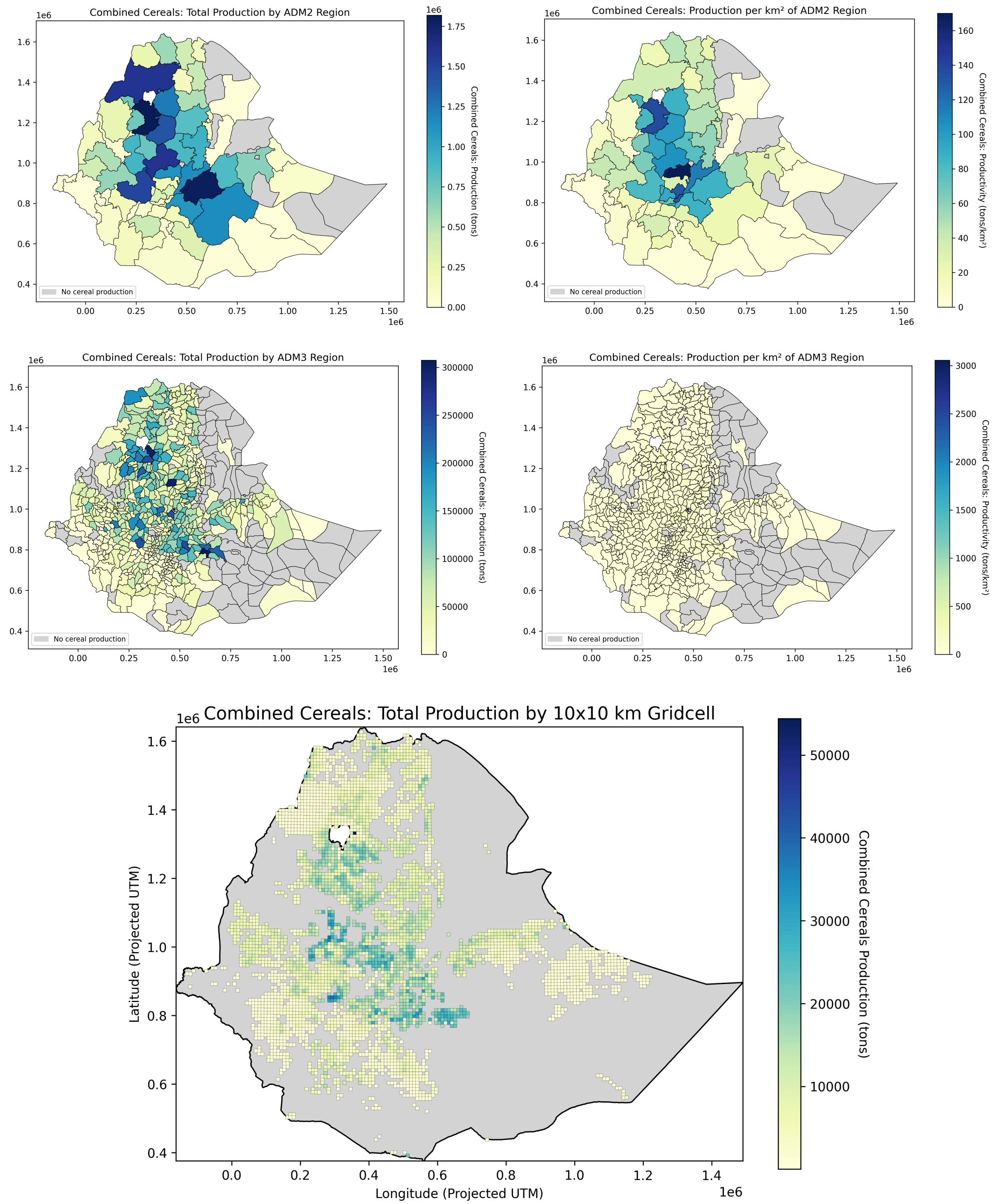


Figure A3: Cereal Crop Production and Productivity by Administrative Region and Gridcell

Data Source: International Food Policy Research Institute (IFPRI), 2024

Wheat

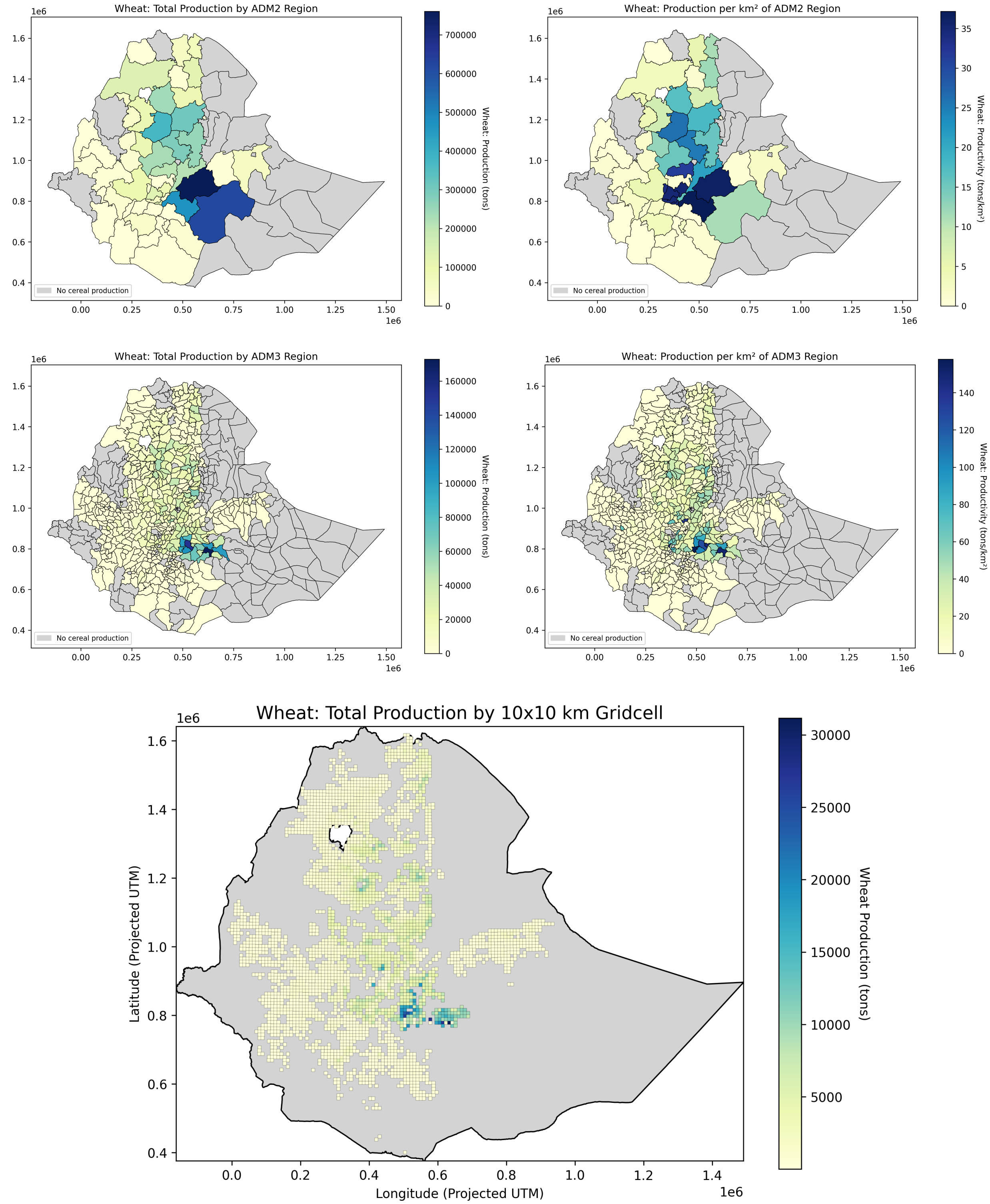


Figure A4: Wheat Production and Productivity by Administrative Region and Gridcell
Data Source: International Food Policy Research Institute (IFPRI), 2024

Rice

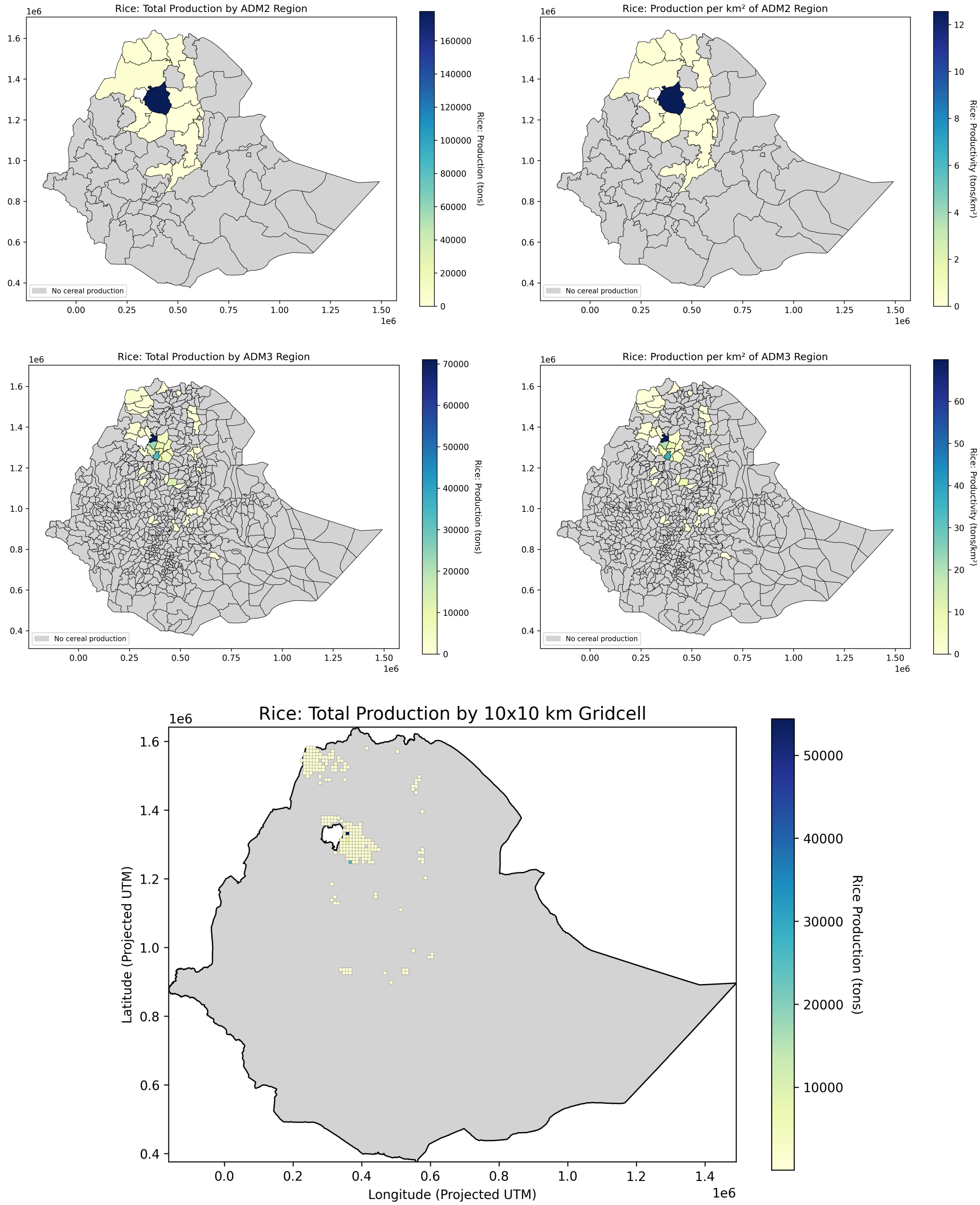


Figure A5: Rice Production and Productivity by Administrative Region and Gridcell
Data Source: International Food Policy Research Institute (IFPRI), 2024

Maize

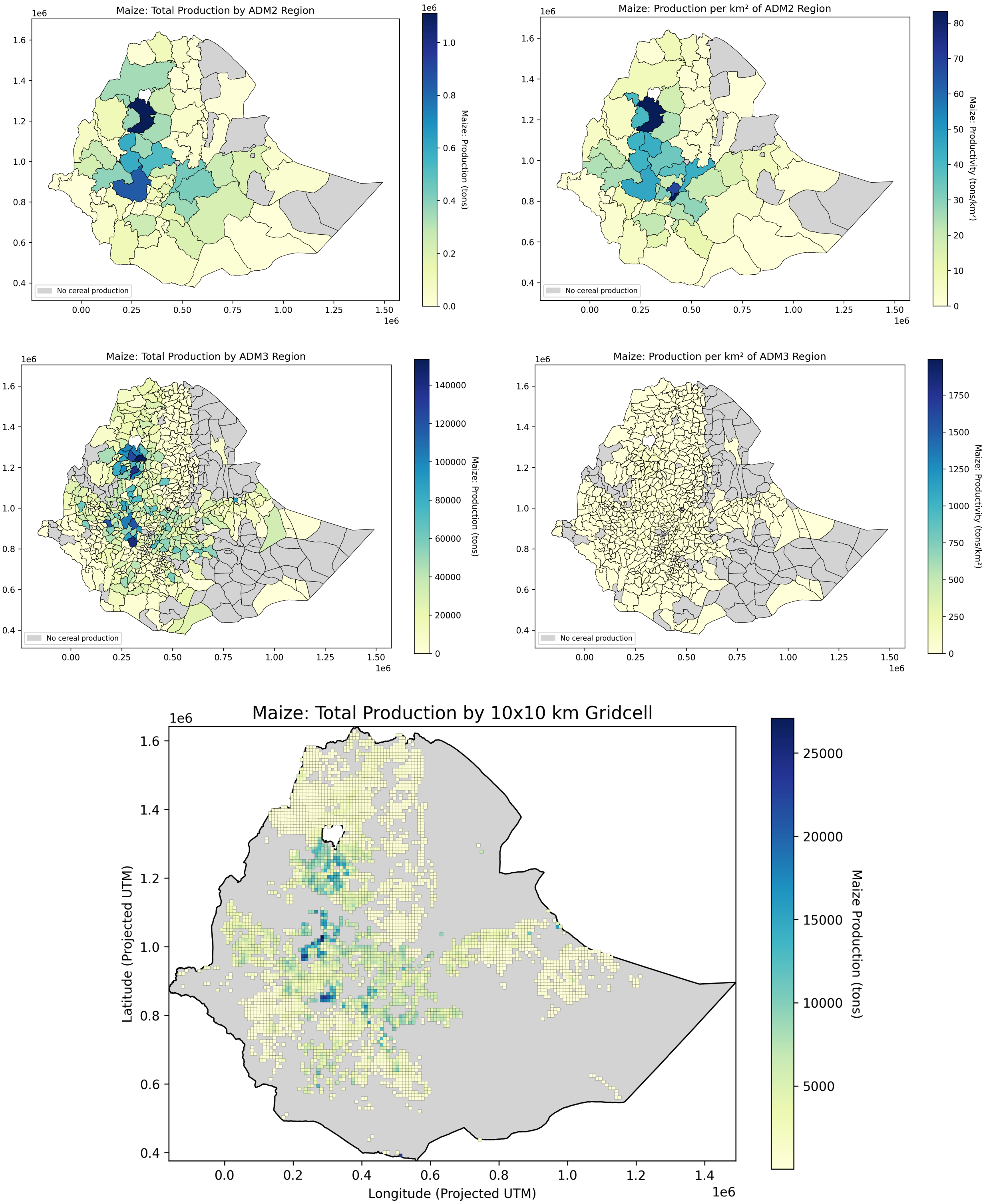


Figure A6: Maize Production and Productivity by Administrative Region and Gridcell
Data Source: International Food Policy Research Institute (IFPRI), 2024

Barley

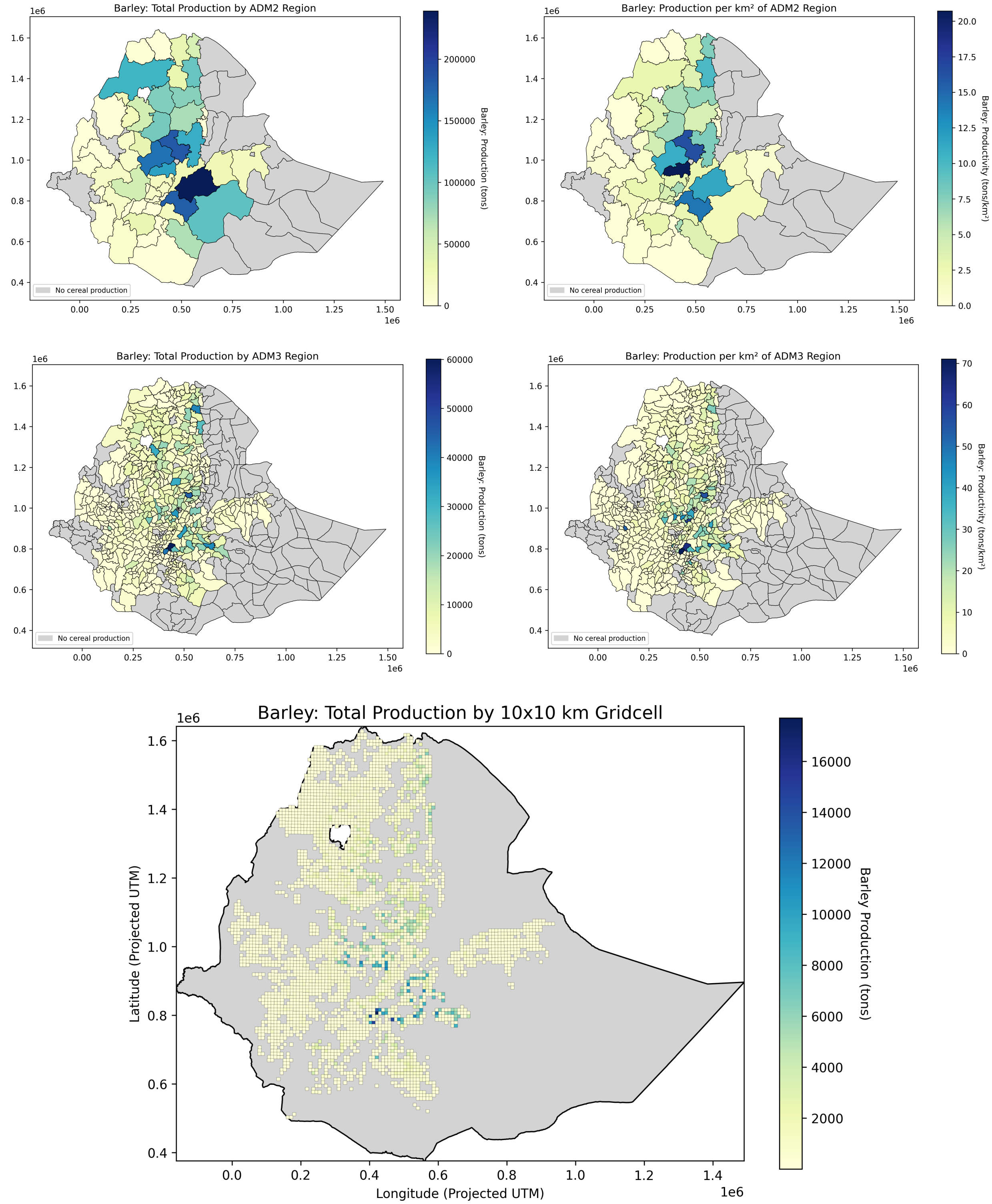


Figure A7: Barley Production and Productivity by Administrative Region and Gridcell
Data Source: International Food Policy Research Institute (IFPRI), 2024

Millet

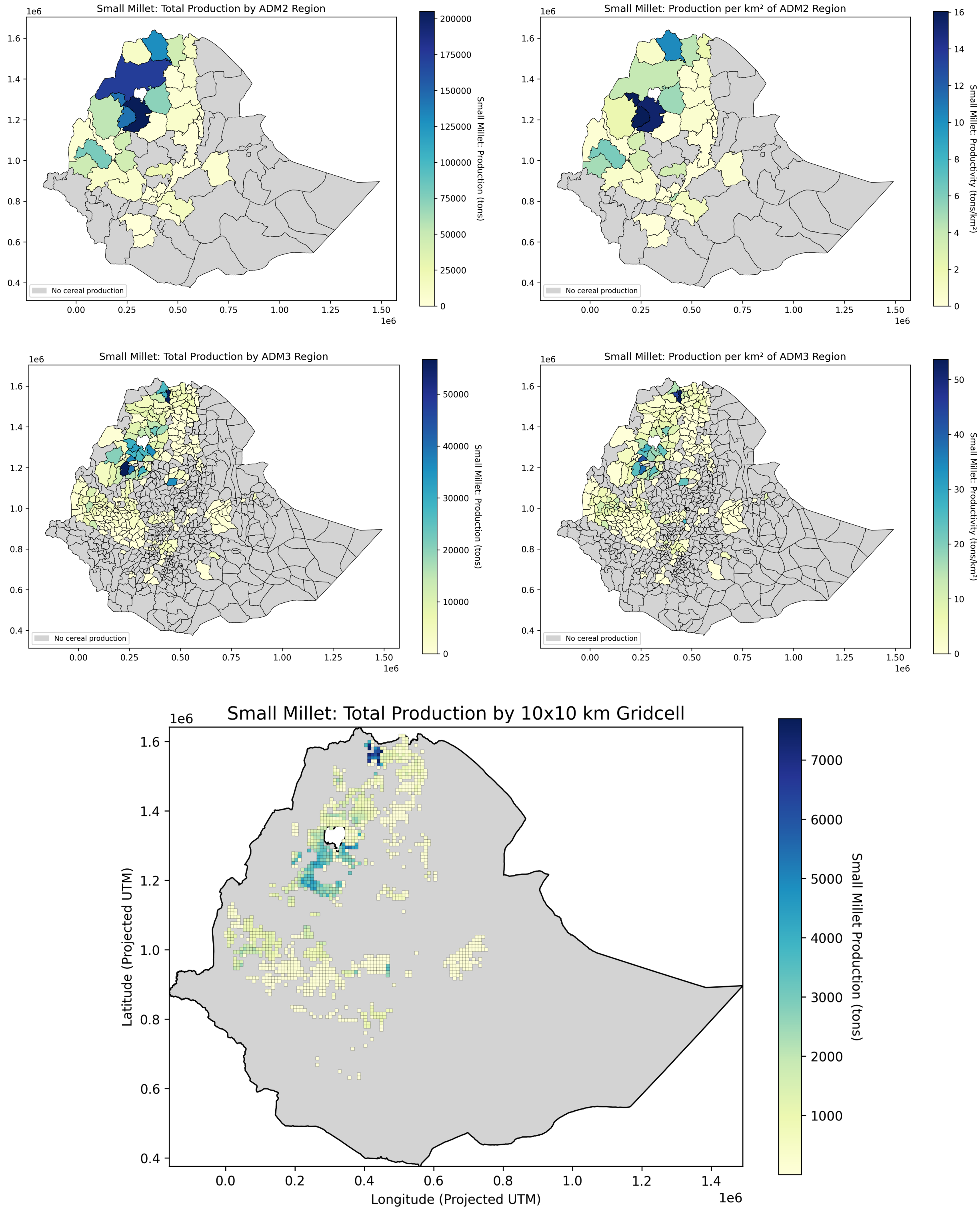


Figure A8: Millet Production and Productivity by Administrative Region and Gridcell

Data Source: International Food Policy Research Institute (IFPRI), 2024

Sorghum

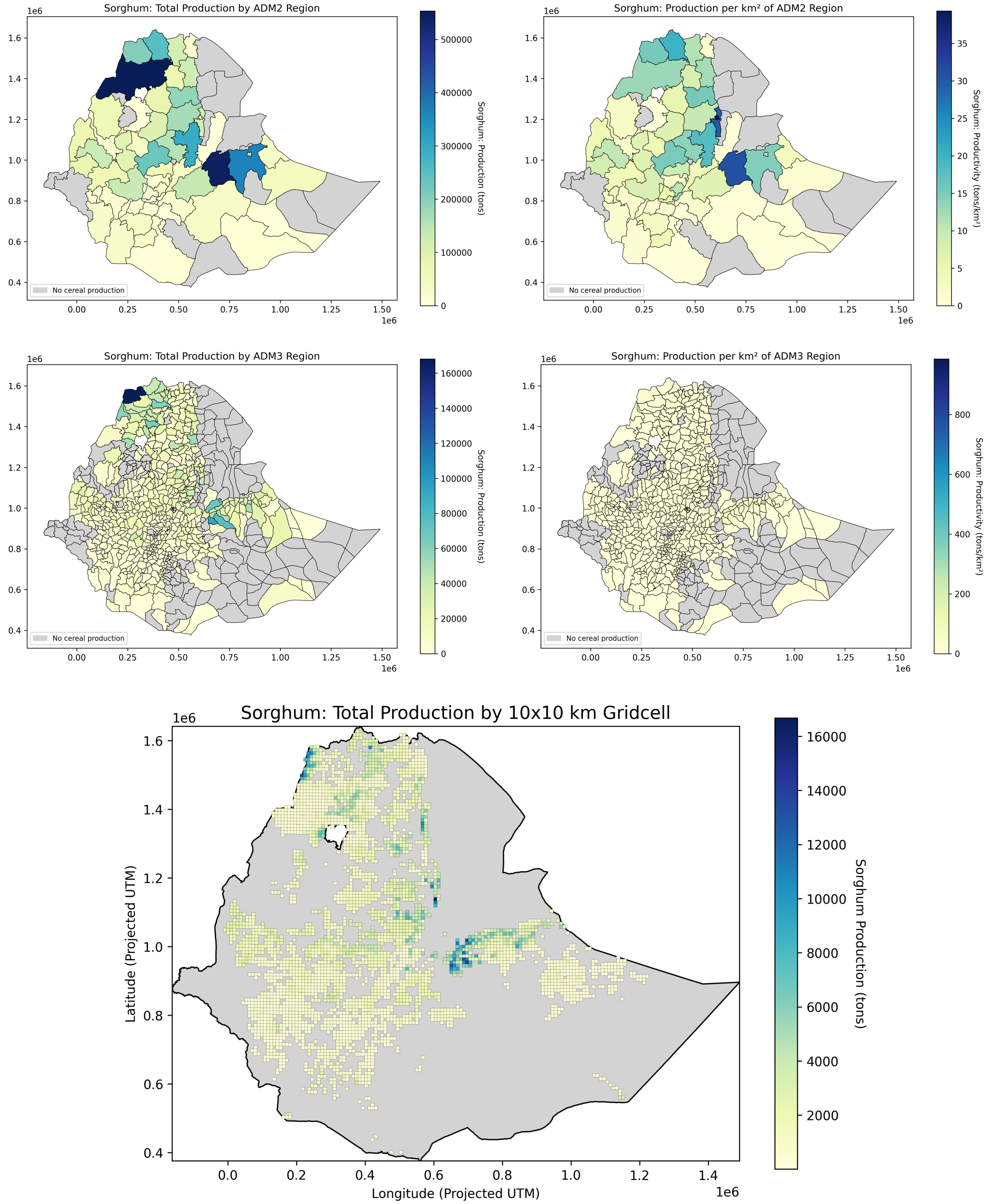


Figure A9: Sorghum Production and Productivity by Administrative Region and Gridcell
Data Source: International Food Policy Research Institute (IFPRI), 2024

Other Cereals

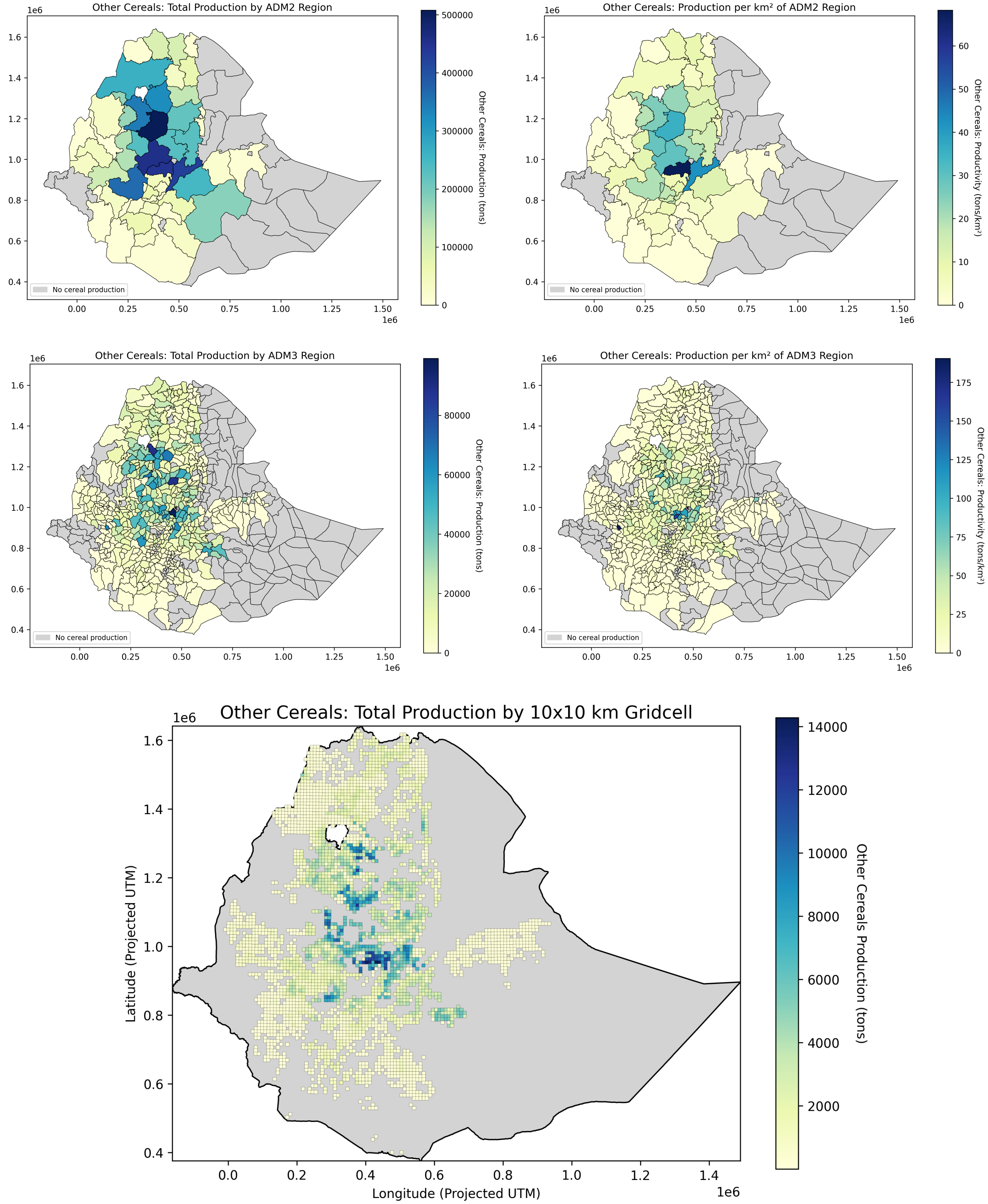


Figure A10: Other Crop Production and Productivity by Administrative Region and Gridcell
Data Source: International Food Policy Research Institute (IFPRI), 2024

Individual Cereal Information

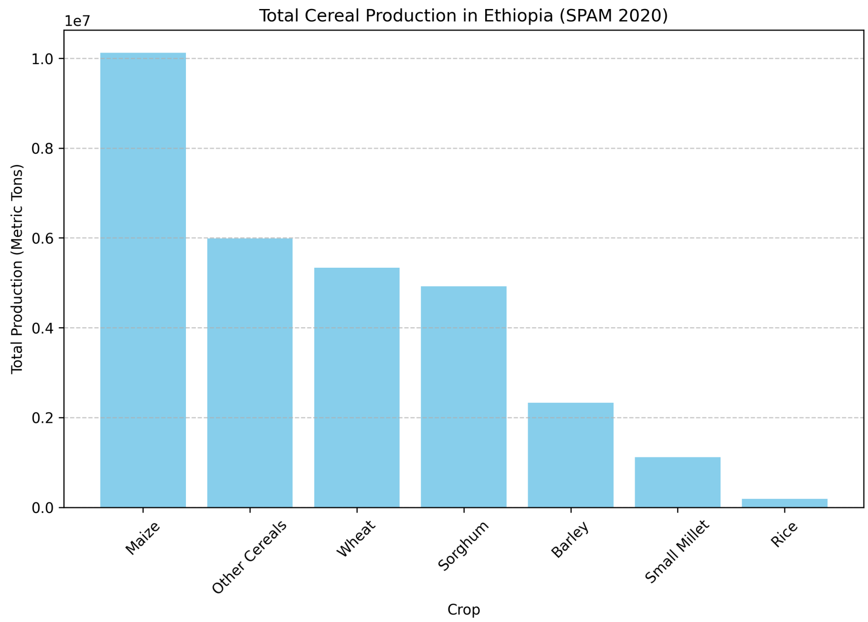


Figure A11: Total annual production per cereal crop in Ethiopia.
Data Source: International Food Policy Research Institute (IFPRI), 2024

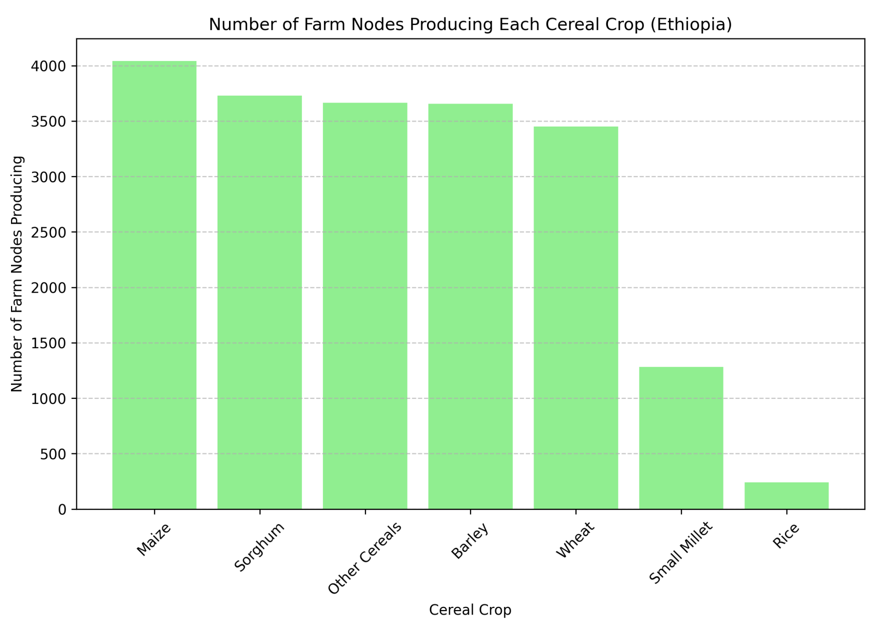


Figure A12: Total 10x10 km gridcells producing each cereal crop.
Data Source: International Food Policy Research Institute (IFPRI), 2024

Dominance Distribution Summary Statistics

Statistic	Dominance Ratio
Count	4316
Mean	0.582
Standard Deviation	0.215
Minimum	0.201
25th Percentile	0.415
Median	0.540
75th Percentile	0.698
Maximum	1.000

Table 4.1: Summary statistics for the crop dominance ratio across farm grid cells.
Data Source: International Food Policy Research Institute (IFPRI), 2024

SPAM Dataset Scope

Crops Included (46)

Wheat, Rice, Maize, Barley, Small Millet, Pearl Millet, Sorghum, Other Cereals, Potato, Sweet Potato, Yams, Cassava, Other Roots, Bean, Chickpea, Cowpea, Pigeon Pea, Lentil, Other Pulses, Soybean, Groundnut, Coconut, Oilpalm, Sunflower, Rapeseed, Sesame Seed, Other Oil Crops, Sugarcane, Sugarbeet, Cotton, Other Fibre Crops, Arabic Coffee, Robust Coffee, Cocoa, Tea, Tobacco, Banana, Plantain, Citrus, Other Tropical Fruit, Temperate Fruit, Tomato, Onion, Other Vegetables, Rubber, Rest Of Crops.

Identifying Fields

Each pixel in the SPAM dataset is characterized by the following fields:

- **grid_code** - Unique pixel identifier
- **ADM0_NAME** - Country name (FIPS0)
- **ADM1_NAME** - First administrative division (FIPS1)

- **ADM2_NAME** - Second administrative division (FIPS2)
- **x** - Longitude of the pixel centroid
- **y** - Latitude of the pixel centroid
- **rec_type** - Variable type (see below)
- **tech_type** - Technology type (see below)
- **unit** - Measurement unit of the variable
- **year_data** - Year of the data. The 2020 version is an average between the years 2019–2021.

Variables: Each crop is associated with four primary variables:

- **A** - Physical area
- **H** - Harvested area
- **P** - Production
- **Y** - Yield

Technologies: Crop production is identified by technology types:

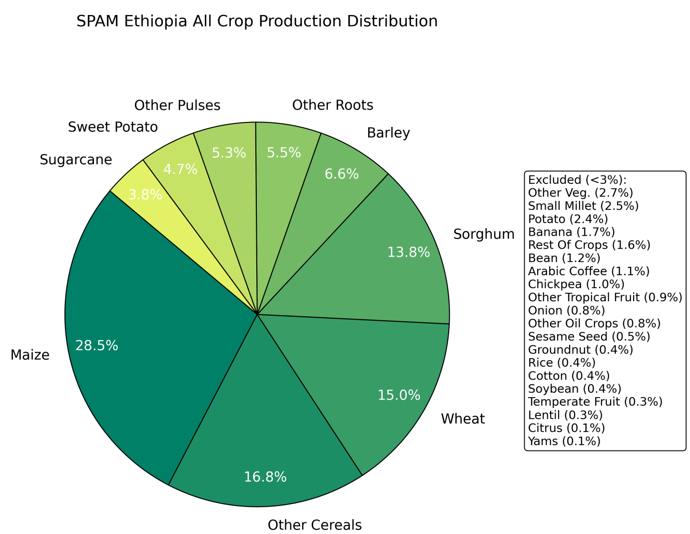
- **TA** - All technologies together
- **TI** - Irrigated portion of the crop
- **TR** - Rainfed portion of the crop (TA - TI)

The cereal crops included in SPAM for Ethiopia are:

- WHEA_A → Wheat
- RICE_A → Rice
- MAIZ_A → Maize

- BARL_A → Barley
- MILL_A → Small Millet
- PMIL_A → Pearl Millet
- SORG_A → Sorghum
- OCER_A → Other Cereals

The "A" represents that this data includes all farming technologies - both irrigated and non-irrigated (International Food Policy Research Institute (IFPRI), 2024).



SPAM Crop Production Distribution for Ethiopia
Data Source: International Food Policy Research Institute (IFPRI), 2024

Bibliography

African Development Bank. (2024). *Ethiopia economic outlook* [Accessed: 2025-03-25]. https://www.afdb.org/en/countries/east-africa/ethiopia/ethiopia-economic-outlook

African Postharvest Losses Information System (APHLIS). (2023, May). Postharvest losses by value chain step in ethiopia 2019 [Accessed: 10 March 2025]. https://www.aphlis.net/en/data/tables/value-chain/ET/all-crops/2019?metric=prc

Alenezy, E. J. (2021). An integer programming formulation of capacitated facility location problem. *International Journal of Mathematics and Computer Science*, *16*(4), 1087–1101.

AmeriGEOSS. (2025). Crop storage final location: Cereal (ethiopia - 1km) [Accessed: January 25, 2025]. https://data.amerigeoss.org/dataset/83c856f2-5b20-4d8b-ab38-e1f8cb4dae79

Bank, T. W. (2018, May). *Cereal market performance in ethiopia: Policy implications for improving investments in maize and wheat value chains*. Agriculture Global Practice, The World Bank Group. Washington, D.C.

Barabási, A.-L. (2016). *Network science* [Available online at https://networksciencebook.com]. Cambridge University Press. https://networksciencebook.com

Berhane, G., Mcbride, L., Tafere, K., & Tamiru, S. (2012, January). Patterns in foodgrain consumption and calorie intake. University of Pennsylvania Press. https://doi.org/10.9783/9780812208610.190

Berhe, M., Subramanyam, B., Chichaybelu, M., Demissie, G., Abay, F., & Harvey, J. (2022). Post-harvest insect pests and their management practices for major food and export crops in east africa: An ethiopian case study. *Insects*, *13*(11). https://doi.org/10.3390/insects13111068

Chandra, S., Sarkhel, M., & Vatsa, A. K. (2021). Capacitated facility location–allocation problem for wastewater treatment in an industrial cluster. *Computers & Operations Research*, *132*, 105338. https://doi.org/https://doi.org/10.1016/j.cor.2021.105338

D'Andreagiovanni, F., Mett, F., & Pulaj, J. (2016). An (mi)lp-based primal heuristic for 3-architecture connected facility location in urban access network design. In G. Squillero & P. Burelli (Eds.), *Applications of evolutionary computation* (pp. 283–298). Springer International Publishing.

Emami, A., Hazrati, R., Delshad, M. M., Pouri, K., Khasraghi, A. S., & Chobar, A. P. (2024). A novel mathematical model for emergency transfer point and facility location. *Mağallaṭ Al-Abḥath al-Handasiyyatˇ*, *12*(1), 182–191. https://doi.org/10.1016/j.jer.2023.11.004

Ethiopian Space Science and Geospatial Institute (SSGI). (2020, April). Addis ababa city administration woreda boundary [Vector Data. Public Domain. Retrieved from Ethionsdi.gov.et].

Food and Agriculture Organization of the United Nations. (2022). Crop storage final location: Cereal (ethiopia - 1km). https://data.apps.fao.org/catalog/dataset/83c856f2-5b20-4d8b-ab38-e1f8cb4dae79

Food and Agriculture Organization of the United Nations (FAO). (2017). *The future of food and agriculture – trends and challenges* [© FAO, 2017]. Food; Agriculture Organization of the United Nations. https://www.fao.org/publications/

Forrest, J., Ralphs, T., Vigerske, S., Santos, H. G., Hafer, L., Kristjansson, B., Fasano, J. P., Straver, E., Jan-Willem, Lubin, M., Lougee, R., Andre, A., Goncalves, J. P., Brito, S., Gassmann, H. I., Cristina, Saltzman, M., Tost, T., Pitrus, B., . . . To-St. (2024). Coin-or/cbc: Release releases/2.10.12. https://doi.org/10.5281/ZENODO.2720283

Gillies, S. (2024a). Rasterio: Geospatial raster i/o for python. https://rasterio.readthedocs.io/

Gillies, S. (2024b). Shapely: Geometric manipulations in python. https://shapely.readthedocs.io/

Gillies, S., Buffat, R., Arnott, J., Taves, M. W., Wurster, K., Snow, A. D., Cochran, M., de Andrade, E. S., & Perry, M. (2024, September 3). *Fiona* (Version 1.10.0) [Available at PyPI: https://pypi.org/project/Fiona]. https://github.com/Toblerity/Fiona

Goddard, J. (2024, October 10). Ethiopia: Conflict and food insecurity 40 years on from the 1984 famine [Accessed: 2025-03-14]. https://lordslibrary.parliament.uk/ethiopia-conflict-and-food-insecurity-40-years-on-from-the-1984-famine/

Goedde, L., Ooko-Ombaka, A., & Pais, G. (2024). Winning in Africa's Agricultural Market. https://www.mckinsey.com/industries/agriculture/our-insights/winning-in-africas-agricultural-market#/

Google. (2025). Google maps platform [Accessed: 2025-04-11].

Gurobi Optimization, LLC. (2024). Gurobi Optimizer Reference Manual. https://www.gurobi.com

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th python in science conference* (pp. 11–15).

Haile, K., Gebre, E., & Workye, A. (2022a). Determinants of market participation among smallholder farmers in southwest ethiopia: Double-hurdle model approach. *Agriculture & Food Security*, *11*, 18. https://doi.org/10.1186/s40066-022-00358-5

Haile, K., Gebre, E., & Workye, A. (2022b). Determinants of market participation among smallholder farmers in southwest ethiopia: Double-hurdle model approach. *Agriculture & Food Security*, *11*(1), 18. https://doi.org/10.1186/s40066-022-00358-5

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hassen, A. A., & Chen, B. (2022, May). Big data analytics for agriculture input supply chain in ethiopia: Supply chain management professionals perspective [Supervised by Assistant Professor Niclas Eberhagen; Examined by Associate Professor Päivi Jokela; Course: 5IK50E, 30 credits; Subject: Information Systems; Department of Informatics]. https://www.lnu.se

Hengsdijk, H., & de Boer, W. J. (2017). Post-harvest management and post-harvest losses of cereals in ethiopia. *Food Security*, *9*(5), 945–958. https://doi.org/10.1007/s12571-017-0714-y

HPS, P. (2025). Ethiopia: Getting around by car [Accessed: 2025-04-12].

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. https://doi.org/10.1109/MCSE.2007.55

International Development Research Centre (IDRC). (2025). Leveraging a systems approach for healthier and sustainable diet policy in ethiopia [Accessed: 2025-03-24]. https://idrc-crdi.ca/en/what-we-do/projects-we-support/project/leveraging-systems-approach-healthier-and-sustainable-diet

International Food Policy Research Institute (IFPRI). (2024). Global Spatially-Disaggregated Crop Production Statistics Data for 2020 Version 1.0. https://doi.org/10.7910/DVN/SWPENT

International Fund for Agricultural Development (IFAD). (n.d.). Country facts: Ethiopia [Accessed: January 26, 2025]. https://www.ifad.org/en/w/countries/ethiopia

International Fund for Agricultural Development (IFAD). (2025). The field report: Why small farms matter [Accessed: January 26, 2025]. https://www.ifad.org/thefieldreport/

Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., . . . Leblanc, F. (2020, July). *Geopandas/geopandas: V0.8.1* (Version v0.8.1). Zenodo. https://doi.org/10.5281/zenodo.3946761

Kumar, D., & Kalita, P. (2017). Reducing postharvest losses during storage of grain crops to strengthen food security in developing countries. *Foods*, *6*(1), 8. https://doi.org/10.3390/foods6010008

Lavassani, K. M., Boyd, Z. M., Movahedi, B., & and, J. V. (2023). Ten-tier and multi-scale supply chain network analysis of medical equipment: Random failure & intelligent attack analysis. *International Journal of Production Research*, *61*(24), 8468–8492. https://doi.org/10.1080/00207543.2022.2152892

Li, T., Ma, J., Huang, Y., Zhang, S., Gao, H., & Xu, C. (2025). Disaster analysis and lessons learned from the july 22, 2024, ethiopian landslide. *Earthquake Research Advances*, 100358. https://doi.org/https://doi.org/10.1016/j.eqrea.2025.100358

Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J., & Schipper, A. M. (2018). Global patterns of current and future road infrastructure. *Environmental Research Letters*, *13*(6), 064006. https://doi.org/10.1088/1748-9326/aabd42

Minten, B., Stifel, D., & Tamru, S. (2012, May). *Structural transformation in ethiopia: Evidence from cereal markets* (ESSP II Working Paper No. 39) (Ethiopia Strategy Support Program II (ESSP II)). International Food Policy Research Institute (IFPRI). Addis Ababa, Ethiopia. http://essp.ifpri.info/

Mitchell, S., O'Sullivan, M., & Dunning, I. (2011, September). Pulp: A linear programming toolkit for python. https://github.com/coin-or/pulp

Mogale, D. G., Kumar, S. K., & Tiwari, M. K. (2020). Green food supply chain design considering risk and post-harvest losses: A case study [Copyright - © Springer Science+Business Media, LLC, part of Springer Nature 2020; Last updated - 2024-11-17; SubjectsTermNotLitGenreText - India]. *Annals of Operations Research*, *295*(1), 257–284. http://libproxy.usc.edu/login?url=https://www.proquest.com/scholarly-journals/green-food-supply-chain-design-considering-risk/docview/2471727908/se-2

Mogale, D., Kumar, M., Kumar, S. K., & Tiwari, M. K. (2018). Grain silo location-allocation problem with dwell time for optimization of food grain supply chain network. *Trans-*

portation Research Part E: Logistics and Transportation Review, *111*, 40–69. https://doi.org/https://doi.org/10.1016/j.tre.2018.01.004

Neto, P. J. S., Datta, D., Steiner, M. T. A., Júnior, O. C., Figueira, J. R., Detro, S. P., & Scarpin, C. T. (2017). A multi-objective genetic algorithm based approach for location of grain silos in paraná state of brazil. *Computers & Industrial Engineering*, *111*, 381–390. https://doi.org/10.1016/j.cie.2017.07.019

pandas development team, T. (2020, February). *Pandas-dev/pandas: Pandas* (Version 2.2.3). Zenodo. https://doi.org/10.5281/zenodo.3509134

Papachristou, M., & M, A. R. (2024). Structural measures of resilience for supply chains [Copyright - © 2024. This work is published under http://creativecommons.org/licenses/by/4.0/ (the "License"). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License; Last updated - 2024-09-06]. http://libproxy.usc.edu/login?url=https://www.proquest.com/working-papers/structural-measures-resilience-supply-chains/docview/2789867576/se-2

Python Software Foundation. (2024). Python: A dynamic, open-source programming language. https://www.python.org/

Rashid, S., Getnet, K., & Lemma, S. (2010, July). *Maize value chain potential in ethiopia: Constraints and opportunities for enhancing the system* (tech. rep.) (Working Paper prepared for the Assessment of Key Areas of Ethiopian Agricultural Sector Project, funded by the Bill & Melinda Gates Foundation). International Food Policy Research Institute (IFPRI). Washington, DC. https://www.ifpri.org/publication/maize-value-chain-potential-ethiopia

Ritchie, H., Rosado, P., & Roser, M. (2023). Agricultural production [https://ourworldindata.org/agricultur production]. *Our World in Data*.

Runfola, D., Contributors, C., Rogers, L., Habib, J., Horn, S., Murphy, S., Miller, D., Day, H., Troup, L., Fornatora, D., Spage, N., Pupkiewicz, K., Roth, M., Rivera, C., Altman, C., Schruer, I., McLaughlin, T., Biddle, R., Ritchey, R., ... Hobbs, L. (2020).

Geoboundaries: A global database of political administrative boundaries. *PLoS ONE*, *15*(4), e0231866. https://doi.org/10.1371/journal.pone.0231866

Sasu, D. D. (2022). Cereal production in africa in 2022, by country [Accessed: 03/10/2025]. https://www.statista.com/statistics/1322348/cereal-production-in-africa-by-country/

Snyder, L. V., & Daskin, M. S. (2005). Reliability models for facility location: The expected failure cost case. *Transportation Science*, *39*(3), 400–416. https://doi.org/10.1287/trsc.1040.0107

SPAM. (2025). Methodology - a look behind spam and what makes it run. [Accessed: January 25, 2025]. https://mapspam.info/methodology/

Strubenhoff, H. (2021, February). *Can agriculture be ethiopia's growth engine?* [Accessed: 2025-03-14]. https://www.brookings.edu/articles/can-agriculture-be-ethiopias-growth-engine/

Teferra, T. F. (2022). The cost of postharvest losses in ethiopia: Economic and food security implications. *Heliyon*, *8*(3). https://doi.org/10.1016/j.heliyon.2022.e09077

Teshome, D. G., Mussa, E. C., Dessie, A. B., Abate, T. M., & Ayalew, D. Y. (2023). Farmers' willingness to pay for metal silo cereal storage technology in northeast ethiopia. *Heliyon*, *9*(8), e18223. https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e18223

Thomas, K., Benjamin, R.-K., Fernando, P., Brian, G., Matthias, B., Jonathan, F., Kyle, K., Jessica, H., Jason, G., Sylvain, C., Paul, I., Damián, A., Safia, A., Carol, W., & Jupyter Development Team. (2016). Jupyter notebooks – a publishing format for reproducible computational workflows. In *Positioning and power in academic publishing: Players, agents and agendas*. IOS Press.

Tiguh, E. E., Delele, M. A., Ali, A. N., Kidanemariam, G., & Fanta, S. W. (2024). Assessment of harvest and postharvest losses of teff (eragrostistef (zucc.)) and methods of loss reduction: A review. *Heliyon*, *10*(9), e30398. https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e30398

Tönissen, D. D., Arts, J. J., & Shen, Z.-J. ((2019). Maintenance location routing for rolling stock under line and fleet planning uncertainty. *Transportation Science*, *53*(5), 1252–1270. https://doi.org/10.1287/trsc.2018.0866

Vincent, D., & Haldane, A. (2024). Contextily: Context geotiles in python [Accessed: 10 March 2025]. https://contextily.readthedocs.io/

Virtanen, P., et al. (2020). SciPy: Open source scientific tools for python. https://doi.org/10.1038/s41592-019-0686-2

Warf, B. (2010). Spatial optimization methods. In *Encyclopedia of geography* (pp. 2658–2659). SAGE Publications, Inc. https://doi.org/10.4135/9781412939591

World Bank. (n.d.-a). Employment in agriculture (% of total employment) (modeled ilo estimate) - ethiopia [Accessed: 2025-03-14].

World Bank. (n.d.-b). Prevalence of severe food insecurity in the population (%) - ethiopia [Accessed: 2025-03-14].

World Food Programme. (2025). *Ethiopia* [Accessed: 2025-03-25]. https://www.wfp.org/countries/ethiopia

WorldData.info. (2025). Speed limits around the world [Accessed: 2025-04-12].

Yibeltal, K. (2023, December). Northern ethiopia facing famine, says tigray official [Accessed: 2025-03-14]. https://www.bbc.com/news/world-africa-67840422