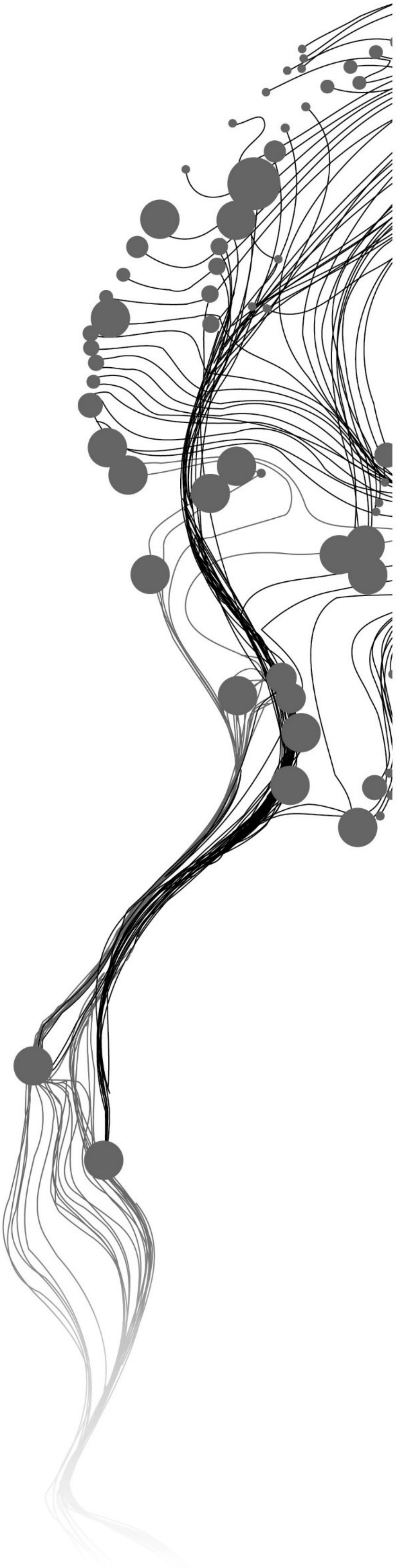
Estimating Consumer Food Accessibility Across Urban-Rural Catchment Areas in Ethiopia

FLORENCIO CAMPOMANES V
June 2022

SUPERVISORS:
Prof. Dr. A. D. Nelson
Dr. M. T. Marshall



Estimating Consumer Food Accessibility Across Urban-Rural Catchment Areas in Ethiopia

FLORENCIO CAMPOMANES V

Enschede, The Netherlands, June 2022

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resources Management

SUPERVISORS:
Prof. Dr. A. D. (Andy) Nelson
Dr. M. T. (Michael) Marshall

THESIS ASSESSMENT BOARD:
Dr. Ir. C.A.J.M. (Kees) de Bie (Chair)
Dr. Ir. B.G.J.S. (Ben) Sonneveld (External Examiner, VU Amsterdam)

ABSTRACT

Food insecurity has long been a problem globally and even more so in low and middle income countries. However, existing indicators to assess and monitor food security are often operationally constrained as these do not provide insights on the underlying causes of food insecurity. One of the under-studied pillars of food security is food access or the ability of people to physically, economically, and socially access food. The food access studies that have incorporated a spatial context have mostly considered physical access but not economic access and have only done so in a limited spatial coverage. Moreover, several social inequalities have been studied but none so far have examined inequalities in food access.

This study aimed to quantify food access and food access inequality across the urban-rural continuum of Ethiopia across time. To do this, this study developed a food access index (FAI) from household survey data using principal component analysis (PCA). Variables related to physical and economic access were found to be majority of the significant variables in the first principal component, but economic access variables had stronger contributions. The newly developed index was predicted across space and time for Ethiopia at high spatial detail using geospatial covariates by selecting important geospatial variables using random forest (RF). A generalized additive model (GAM) was trained to predict the FAI using 4 selected variables, namely, (1) road proximity, (2) travel time to markets, (3) cereal prices, and (4) precipitation at coldest quarter, and achieved an R^2 of 0.62. The predicted FAI across Ethiopia's urban-rural continuum showed that peri-urban areas, which make up a considerable proportion of the population, had low food access. Peri-urban areas of intermediate and small cities had the most temporally volatile food access within a year leaving such areas vulnerable to food price spikes. Finally, food access inequality was quantified by calculating the Gini index from the spatial predictions of the FAI and was found to be low for all of Ethiopia, driven by the large area percentage of small cities and catchment areas where food access inequality was low. However, food access inequality was high in large and intermediate cities and their catchment areas.

From these findings, this study provided a proof-of-concept of the possibilities for spatiotemporal quantification of food access at high spatial detail. Moreover, a scenario development application[1] was also built to simulate possible shocks like price spikes which comes at a time when global food access is threatened by rising prices caused by the war in Ukraine, among others. This study recommends further testing and improving this approach by (1) testing other modelling techniques and (2) implementing at multi-country or global scale to develop more insights into food access across the urban-rural continuum.

Keywords: Food access, inequality, household survey, geospatial data, urban-rural catchment areas, principal component analysis, random forest, generalized additive model, Gini index

[1] Available at bit.ly/FoodAccessScenarios

ACKNOWLEDGEMENTS

This document is not merely a thesis, but a culmination of constant hard work, failures, and successes, and of course, the support and love from everyone around me.

When I started this journey called the MSc thesis phase, I did not know what to expect or if I was going to achieve anything from this. I was extremely blessed to have two of the most amazing people as my supervisors: Prof. Andy Nelson and Dr. Michael Marshall, who, throughout this process, made me feel like what I was doing made sense and were always honest with me when I wasn't making sense. Andy sent me an article from which we slowly developed the idea of this thesis, and he has only shaped me to become a better scientist with his critical and exceptional mind since I started working with him. Michael was honest with me in the beginning that this topic was not necessarily his expertise but still, he helped me a ton with the technicalities of my methods, especially GAMs (and yes, I am a GAM convert now). I'm also indebted for the support from the both of you and Dr. Mariana Belgiu during the TGS Award application, regardless of the outcome, as you all saw my potential when I couldn't. Mariana, whose Advanced Image Analysis course was the best I've taken in ITC, always gave her never-failing support and endless rays of positivity which always had me leaving our conversations with a smile and a little bit more confidence in myself. I felt like the three of you treated me as family, and I am forever thankful.

This journey wouldn't have been as fun, productive, and healthy (planks and squats!) if not for my Studio 5-036 family. You have all been there through my ups and downs and I cannot thank you enough. Sahara, my constant confidante and sanity check during the last two years with our deep talks and laughs. Biman, video-making brother from another mother and fellow tech geek. Lynette, rickrolling co-raccoon enthusiast and other-planet discoverer. Sachita, the seriously organized friend that somehow still finds time for running. Diana, super caring and bike-riding friend who has great taste in party music. Binita, awesome cook (momo!) and savior of the elephants. Nithesh, academic "stalker" and source of all future opportunities. Our friendship has been bound by our struggles and successes and will withstand any distance and oceans.

At the start of the second year, I unexpectedly made such cool friends and I'm grateful for you and our "therapy sessions" and fruit nights. To Xuanya, Archi, and Cham, for accepting the oversharing and talkative me and just being the special people that you are. To Aulia (the universe), for giving me consent to finish this MSc degree. To Art, salamat kay naa tawn kapwa-Bisaya sa ITC, naa ko'y ma-relate-an diri.

To THE Lace Puno and Kim who've been the best support system outside the Netherlands. We've finally made it! Thank you for listening to my rants and long discussions about my life here in NL and my thesis even if I get all nerdy about it. I appreciate all that you've done and I hope I've made you proud.

To the dad up above, I just know you've been behind me all this time pushing me to achieve excellence. Since the beginning, I've kept my focus on the goal as both you and mom have raised me to do. The Puno-Campomanes spirit is definitely in me and I hope I've somehow made you smile and proud up there.

After almost a year of working on this thesis, I've only given my utmost laser focus in delivering the best I can. And with everyone's love and support, I was able to muster up the confidence to really believe in myself. And finally, here I am, writing the last and most difficult section to write in the MSc thesis: the acknowledgement section, still focused as hell.

"Rise up. Laser focus." – Tom Brady

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC	Akaike information criterion
ANOVA	Analysis of variance
CPI	Consumer price index
CPS	Crop production system
CRP	Cereal price
DHS	Demographic and Health Surveys
ESA	European Space Agency
FABDEM	Forest and buildings removed Copernicus digital elevation model
FAI	Food access index
FAO	Food and Agriculture Organization
FEWSNET	Famine early warning systems network
FIES	Food insecurity experience scale
GAM	Generalized additive model
GCV	General cross validation
GLM	Generalized linear model
HFIAS	Household food insecurity access scale
HH	Household
IFAD	Intenational Fund for Agricultural Development
IIASA	International Institute for Applied Systems Analysis
IPC	Integrated phase classification
LMIC	Low- to middle- income countries
LSMS	Living standards measurement study
MAE	Mean absolute error
MapSPAM	Map spatial production allocation model
MLR	Multilinear regression
NDVI	Normalized difference vegetation index
nMAE	Normalized mean absolute error
nRMSE	Normalized root mean square error
OOB	Out of bag
OSM	OpenStreetMap
PCA	Principal component analysis
PDP	Partial dependence plot
PIRLS	Penalized iterative reweighted least squares
PPP	Purchasing power parity
RF	Random forest
RMSE	Root mean square error
RPX	Road proximity
SDG	Sustainable Development Goals
SSA	Sub-Saharan Africa
TTM	Travel time to markets
URCA	Urban-rural catchment areas
VSURF	Variable selection using random forest
UN	United Nations
WFP VAM	World Food Programme Vulnerability Analysis and Mapping

1. INTRODUCTION

1.1. Background

1.1.1. Food insecurity

The long term trend of declining global hunger has recently reversed, with an estimated 821 million people estimated to be undernourished (FAO et al., 2018). This increase is even accelerating in sub-Saharan Africa (SSA) where population projections have increased from 1.52 billion in 1998 to 2.12 billion in 2019 (Ezeh et al., 2020) with 24.1% suffering from undernourishment in 2020 (FAO et al., 2021). This means that the severity of food insecurity in the region is increasing and the need for evidence-based solutions to reduce and eventually eradicate food insecurity is critical. Food security is defined as "when all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for an active and healthy life" (Campbell, 1991, pp. 408-409; FAO, 1996). This definition incorporates the four pillars of food security: availability, utilization, access, and stability. Food availability relates to food production, import, and stocks while utilization focuses on the quality and nutritional value of food. Food access focuses on distribution, market infrastructure, and affordability. Lastly, stability means the upkeep of the other three pillars of food security over time, even after periodic food insecurity or shocks (e.g. conflict or climatic crisis).

A standard global framework used in crafting solutions towards food security is the United Nations' Sustainable Development Goal 2 (SDG 2) which specifically aims to "end hunger, achieve food security and improve nutrition and promote sustainable agriculture" (UN General Assembly, 2015, p. 14). Directly related to food security, the first target of SDG 2 (SDG 2.1) is to "end hunger and ensure access by all people" (UN General Assembly, 2015, p. 15) by 2030. In order to properly monitor the progress towards such targets, indicators for each target were proposed. SDG indicator 2.1.2 utilizes the Food Insecurity Experience Scale (FIES) (Cafiero et al., 2016) to evaluate the prevalence of food insecurity (United Nations, 2017). This is the only indicator under SDG 2.1 that explicitly and directly looks at food security since the other indicator, prevalence of undernourishment, is a manifestation of either lack of access to food, in general, or lack of access to healthy food. The FIES ranks the severity of food insecurity from (lowest severity) worrying about not acquiring food, consuming food of lower quality, eating food in less quantity, to (highest severity) actually experiencing hunger (Graham et al., 2014).

The FIES, however, is only a direct assessment of the status of food insecurity and does not consider the linkages and causalities leading to such conditions. The view of investigating the interrelated components of food security is also in agreement with a proposed transformation to the monitoring framework of food systems that intends to look at different yet relevant themes (Fanzo et al., 2021). Solving food insecurity needs a multi-stakeholder approach where governments alongside non-profit organizations rely on correct quantitative information on not only the state of food security, but also its underlying factors, to determine what kind of humanitarian aid and investments to provide and where to focus such interventions (Elbers et al., 2007).

A large number of studies used a geospatial and remote sensing approach to better understand food production and availability (Foley et al., 2011; Husak et al., 2008; Mohammed et al., 2020; Wu et al., 2018). Many of these have focused on improved crop health and yields (Karthikeyan et al., 2020; Mutanga et al., 2017; Sishodia et al., 2020). Much less attention has been paid to geospatial approaches that can model disparities in food distribution and access despite recognition that understanding the spatial variation of such disparities is vital for food security assessments and local policy planning (Brown, 2016; Lele et al., 2016).

1.1.2. Food access

Food accessibility refers to the physical, economic, and social ability of people to acquire food. The physical aspect of food access can be simplified into transport options, ease of mobility, and the proximity and travel time to food-related services like markets (Lê et al., 2015). Economic food access is the ability to purchase food which is usually driven by a combination of food market prices, household income, and access to financial and social support (Lele et al., 2016) while the social aspect of food accessibility pertains to the social and cultural norms of purchasing certain food items. The importance of food accessibility is reinforced by the idea that food availability alone does not assure food security (Sen, 1981). This is also supported by the simultaneous occurrence of a 9% increase in the undernourished population even with global food production increasing by 12% since 1990 (Barrett, 2010). Furthermore, increases in access to food partially contributed to higher diet diversity in China (Liu et al., 2014; Wang et al., 2017). Frelat et al. (2016) also recommended that increasing agricultural production and reducing yield gaps be complemented with improving food and market access to boost food security in the SSA, and further called for policy harmonization across multiple sectors. The need to accurately measure and map the variations and possible inequalities of food accessibility is crucial for such policy development and can be done using both qualitative and quantitative household (HH) surveys on food access and nutrition (Lele et al., 2016).

Studies on food accessibility have been conducted with both spatial and non-spatial approaches. Non-spatial approaches mostly use surveys that cover a small area (one village or town) over a relatively short period of time (a few days to a few weeks) providing detailed assessments of the complexities of food access. For example, in Cuba, food diaries were used to document the entire journey of staple food like rice and beans from their source and how these reach the consumer (Bono and Finn, 2016). These food diaries also provided insight on how social and informal activities like gift-exchange impact food access, something that the common market-trade perspective usually does not see. Food shortages in rural households in Bangladesh and Ethiopia were modelled using a binary choice model with wealth, shocks (such as drought, flood, and storage loss) and social norms as factors, and found that in Bangladesh, cultural traditions like a dowry negatively impacted household purchasing power and subsequently affected economic access to food (Velazco and Ballester, 2015). The same study found that in Ethiopia, experiencing such shocks as well as participating as consumers in food markets increased the probability of facing food shortages in rural households. Moreover, using the household food insecurity access scale (HFIAS) in the Yayu area in the southwest region of Ethiopia, differing levels of food insecurity were reported in 53% of households (Usman and Callo-Concha, 2021). The importance of access to food markets in Ethiopia was reinforced by Usman and Haile (2019) where an increase in the distance to the nearest market meant a substantial decrease in food consumption expenditure. While these studies provide a deep dive into food access, they mostly miss two key elements which are physical access and spatial context.

Previous studies have quantified food accessibility using various spatial accessibility measures with different distance calculations with results showing no clear pattern in terms of underlying factors affecting food access. For example, proximity to the nearest market using network distance was used to assess the level of food access across three different time periods and found low levels of food access in areas with low income and minority neighbourhoods (Kolak et al., 2018). This was consistent with findings from Garcia et al. (2020) who employed a population-adjusted kernel density estimation in relation to food stores to calculate food access. Furthermore, to adapt solutions in a food context, the two-step floating catchment area (2SFCA) method (Radke and Mu, 2009) was modified to consider both supply and demand and this also yielded similar results where wealthy areas had higher levels of food access (Chen, 2017). All these coincide with the concept of "deprivation amplification" which says that economically challenged areas tend to have lower access to resources (Macintyre, 2007). However, another study that used the number of grocery stores within a geographic boundary to represent market density, reported conflicting findings where areas with more grocery stores had higher poverty rates (Jeong and Liu, 2020). A spatial lag model produced matching results where economic deprivation was associated with higher food access (Jin and Lu, 2021).

This disagreement in results suggests that solely relying on physical access to represent access may be inadequate because even with good physical access, maximizing the use of the transport infrastructure may not be economically viable (Nelson et al., 2019). In the context of food access, travel costs to the nearest market or the food being sold could be unaffordable. Thus, for a more accurate representation of food accessibility, the inclusion of economic access to food in its calculation is a must. There have been studies that considered both economic and physical access to food (Breyer and Voss-Andreae, 2013; Jiao et al., 2012) where food affordability was emphasized but these were conducted at local scale, either for a single city or town. However, both physical and economic access to food could vary substantially across urban, peri-urban, and rural areas (Janda et al., 2021; Nyangasa et al., 2019).

1.1.3. Urban-rural disparities

To create a clear spatial representation of the urban-rural systems, urban-rural catchment areas (URCAs) were mapped by Cattaneo et al. (2021b) at a global scale. They observed that only about 20% of the global population live in or move around large urban cities (population > 1 mil.) while more than 50% are located in small cities (population < 0.25 mil.) and its surroundings. Furthermore, the same study showed the importance of peri-urban areas of intermediate (population of 0.25-1 mil.) and small cities since about 25% of the population in low-to-middle income countries (LMICs) live in such areas. However, there was a scarcity of studies examining the interrelatedness of food systems across urban and rural areas (Abu Hatab et al., 2019). Cattaneo et al. (2021a) further argued that to properly deal with issues related to food, policies and their accompanying studies must be done at larger scales that include urban, peri-urban, and rural areas, which the URCAs provide. Furthermore, there are limited studies on inequalities related to food systems from a spatial perspective at both intra-urban and inter-country scale but none across the urban-rural continuum.

Inequalities in undernutrition and stunting, both health manifestations of lack of food access, have grown since 1975 (Bell et al., 2021). The same study found the concentration of such health issues mainly in LMICs like India, Indonesia, and Ethiopia. Further, analyses of inequalities in food availability and nutritional value should be accompanied with access, both physical and economic, as excluding access may only underestimate inequality given that prices of food high in nutrition can largely vary (Headey and Alderman, 2019). While food access inequalities have been studied (Ball et al., 2009; Li et al., 2019), none have done so at larger scales, like at country-level. And to fully grasp the extent of inequality in our food systems, large-scale, granular, and subnational data are required to pinpoint various demands and disparities in food system measures (Downs and Fox, 2021).

There have been very few large-scale (regional to global) studies on food access. HH survey data from 15 countries in SSA were analyzed and found child stunting and undernutrition, health outcomes associated with food insecurity (Gundersen and Ziliak, 2015), to be worse in rural areas than urban areas (Fotso, 2007). In the United States, county-level food access, measured as access to different types of stores and markets, was calculated for all non-metropolitan counties in the country (Deller et al., 2015). Pozzi and Robinson (2008) quantified food access as travel time to the nearest market in seven countries in East Africa. To include informal markets, their definition of a market was extended to thinly populated urban areas but also acknowledged that their database was still incomplete in terms of location of markets and urban centers. A possible reason for the lack of large-scale analyses is the unavailability or incompleteness of input data. For example, road networks or market locations outside large cities are either incomplete or not detailed enough. To partly address these data gaps, many researchers and organizations have recently made their datasets and results openly available.

1.1.4. Recent advancements in data availability and analysis techniques

In 2015, a global friction surface dataset at 1×1 km resolution, sourced from OpenStreetMap (OSM) and Google, that estimated the number of minutes it takes to travel one meter of the Earth's surface, was made openly available (Weiss et al., 2018). Travel time to cities was estimated on the basis of this friction surface and its relevance was demonstrated through correlations with socioeconomic outcomes at a global scale (Weiss et al., 2018). This was further expanded by Nelson et al. (2019) by calculating the travel time to cities of different population sizes and found high levels of spatial variability specifically in sub-Saharan Africa. The previously mentioned URCAs (Cattaneo et al., 2021b) were also derived using an updated version of the friction surface dataset (Weiss et al., 2020). This gave a clearer connected picture, in terms of access to services, of not only rural and urban areas but also the peri-urban areas that usually go unnoticed. All these recently developed datasets and results pave the way for new accessibility related studies.

Another data source is the Living Standards Measurement Study (LSMS) by The World Bank which has collected HH surveys in several countries, including lower-middle income countries (LMICs), to better understand welfare and behaviour and consequently craft local policies (Grosh et al., 1998). It has been used in many different applications in several LMICs like assessing poverty (Menchini and Redmond, 2009), education (Lohani et al., 2010), and food security (Zezza and Tasciotti, 2010). Another source of HH level survey data is the Demographic and Health Surveys (DHS), a similarly structured dataset to the LSMS, that focuses more on health and nutrition but also provides some economic insights.

Indices have been developed to consolidate and quantify the responses from the numerous questions in these surveys. For example, an asset-based wealth index was constructed from the DHS data to quantify wealth by applying principal component analysis (PCA) and using the first principal component to explain wealth (Filmer and Pritchett, 2001; Rutstein and Johnson, 2004). Other methods for single index aggregation have also been used. For example, the Human Development Index uses the geometric mean of several country-level indicators like education, life expectancy, and standard of living (UNDP, 2020). These calculated indices from HH surveys provide a direct quantifiable measure of different aspects of living and present opportunities for more quantitative studies.

An important aspect of these surveys is that most are geolocated which means they provide spatial information about food insecurity, access to services, wealth, and food prices at household and community level. To exploit the spatial aspect of the data, Jean et al. (2016) integrated the expenditure data from the LSMS data and the wealth index from DHS with nightlight and other satellite imagery to estimate village-level expenditure and poverty in five countries, including remote areas where surveys have not yet been conducted. Also, since satellite imagery has been available for more than a decade, Yeh et al. (2020) successfully trained a deep learning model on multispectral imagery to estimate cluster- and district-level economic well-being in 23 countries in Africa over time. Recently, Chi et al. (2022) improved upon this and estimated relative wealth from the DHS datasets but using finer spatial grids of 2.4 km resolution instead of spatially-aggregated estimates and at a larger spatial coverage of 135 LMICs across the world.

Several approaches to national scale analyses for food security have been tested, although not specifically in a food access context. Lentz et al. (2019) used HH surveys and geospatial data like demography, rainfall, markets, and geography in linear and log-linear regressions to predict three food security measures for village clusters for the entirety of Malawi, at 83 to 99% accuracy. In a similar approach, food security transitions, specifically the change in Integrated Phase Classification, for all livelihood zones in Ethiopia were forecasted from geospatial covariates and HH surveys using machine learning and achieved promising results in long term forecasting (Westerveld et al., 2021). Both studies showed the potential of HH surveys and geospatial covariates in tandem with data-driven approaches like machine learning to conduct reliable national scale food security assessments. These approaches, combined with the availability of the friction surface dataset, its derivative works like the URCA classification, and the array of LSMS and DHS surveys, opens possibilities for novel accessibility studies because of their global spatial coverage and the use of such datasets together in the context of food security has not yet been done at large scales.

1.2. Problem statement

Food insecurity has been a long-standing problem in SSA. Indicators to assess food insecurity status in the SDG framework are limited and do not give much insight to the individual pillars of food security and the underlying factors. Many studies have investigated food production and food availability at both small and large scales. However, food availability is only one pillar of food security and while several studies have also been conducted to investigate and monitor food access, most of these were either done at local scale (e.g. village or town level) or were non-spatial in nature. Those studies fail to describe the disparity across urban, peri-urban, and rural areas.

Clearly, there is a lack of research in terms of investigating the differences in food access across urban and rural food systems (Abu Hatab et al., 2019). Policies to account for differences in education and health across the urban-rural continuum have been proposed but not for food systems (Cattaneo et al., 2022). Only one study has focused on food accessibility in parts of SSA at a large scale (e.g. country or multi-country level) from a geospatial perspective (Pozzi and Robinson, 2008) but had considerable data limitations. Another shortcoming of existing studies, including Pozzi and Robinson (2008), is that these have not considered both physical and economic access side-by-side as most have only assessed food access in terms of physical access. The studies that considered both economic and physical access were (1) intra-urban analyses and (2) only included affordability, in terms of supermarket prices, and not purchasing power or wealth (Breyer and Voss-Andreae, 2013; Jiao et al., 2012). Similarly, studies that examined inequalities in food access from a spatial context have not done so at country-scale (Ball et al., 2009; Li et al., 2019). Lastly, even if we are in a time when so many different datasets are available, like the LSMS survey data in Ethiopia, along with the fast-paced development of various techniques in the field of artificial intelligence and spatial analysis, there is still a lack of granular and disaggregated (not bound by administrative boundaries) information on food access when detailed targeting of interventions and policy crafting is needed.

1.3. Research objectives and questions

This study primarily aims to combine the use of HH survey and geospatial data to estimate food accessibility, accounting for both physical and economic access, across URCAs in Ethiopia and then to quantify the inequality in food access from these estimations. To do so, the study intends to develop a food accessibility index for Ethiopia at a high spatial detail. Such a food accessibility index can be used together with the FIES to better understand the underlying causes of food insecurity, specifically in terms of food access.

The detailed sub-objectives, research questions, and hypotheses are:

a) To construct a **food accessibility index** from LSMS survey data using principal component analysis.
   1. Which LSMS indicator variables align together in each principal component?
   Hypothesis: Economic and physical access will be of equal importance.

b) To estimate an intra-annual **disaggregated food accessibility index** for the entire Ethiopia using geospatial data.

1. What geospatial factors significantly contribute to the food accessibility index in Ethiopia?
Hypothesis: All geospatial factors equally contribute to the food accessibility index.
2. What is the difference in levels of the food accessibility index across URCAs?
Hypothesis: The food access index does not change with distance from urban centers.
3. How does the food accessibility index change across months in a year?
Hypothesis: The food access index is constant across the year.

c) To quantify the level of **food access inequality** across URCAs in Ethiopia.
1. What is the difference in inequality in food accessibility across the different URCAs?
Hypothesis: Food access inequality does not change with distance from urban centers.

1.4. Conceptual diagram

The conceptual diagram takes the entire country of Ethiopia as the encompassing system. It shows that consumers and producers both make use of the transport infrastructure in relation to their respective uses of food and crop markets. Crops, the production of which is dependent on climate, weather, soil characteristics, and farming practices, are sold by the producers/farmers to markets as part of their livelihood. Crop production also affects food prices and in the same way, anything that affects crop production like weather also indirectly impacts food prices. On the other hand, consumers purchase crops according to their purchasing power and asset ownership. Assistance, in any form, be it financial or in-kind, from many different organizations or governments also provides support to communities. Lastly, shocks like conflict, hamper the transport infrastructure limiting physical access to markets.

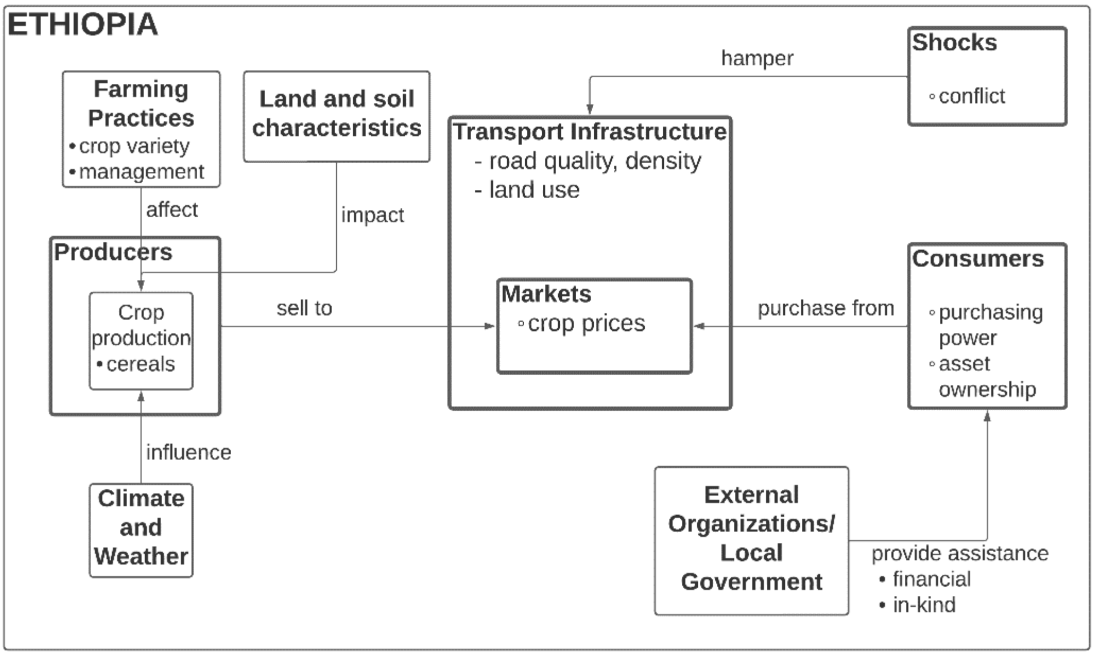


Figure 1. Conceptual Diagram

2. STUDY AREA AND DATASETS

2.1. Ethiopia

The study area is the entire country of Ethiopia which is located in East Africa, specifically in the "Horn of Africa". It is the tenth-largest country in Africa with a surface area of more than 1.1m km² (UNEP, 2008) and a population of almost 100m in 2015. The urban population is expected to double by 2050 (FAO, 2017). The majority of the population reside in the Ethiopian Highlands which is divided by the Great Rift Valley. The majority of agricultural production in the country comes from these highlands which subsequently cause extensive soil erosion and desertification (UNEP, 2008). A large portion of the land in Ethiopia is considered moderately to severely degraded caused by many factors like overgrazing, deforestation, and poor farming practices.

Figure 2. Map of Study Area. *Base map source: ESRI*

As seen in Figure 3, Ethiopia's crop calendar has two seasons: the Belg season, or the short rainy season, which runs from February to late April or May and the Meher season, or the main season, which runs from

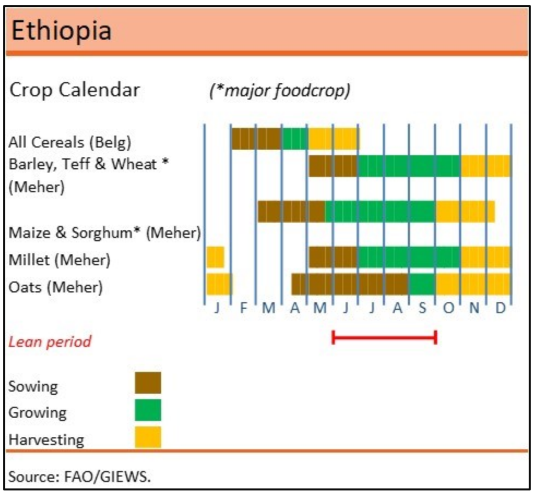


Figure 3. Ethiopia crop calendar. Source: (FAO, 2021b)

around mid-May to September, depending on the area. The middle of the Meher season is usually considered a "lean period" as crops are still growing and there is less available products in the market.

2.2. Living standards measurement study 2018 data

The 2018 LSMS dataset in Ethiopia (Central Statistics Agency of Ethiopia, 2020), also known as the Ethiopia Socioeconomic Survey 2018-19 or ESS4, was the fourth wave of socioeconomic surveys by World Bank and LSMS in Ethiopia, with previous ones being conducted in 2011-12, 2013-14, and 2015-16. The data was collected between September 2018 and August 2019 with separate visits during September-December 2018, February-March 2019, and finally June-August 2019. The data covers 6,804 households in 535 communities which were regionally and nationally representative of urban and rural areas. The data contains several sections covering different themes, some of which were considered more relevant to food access than others. Here, our focus was on variables like access to basic services which related to physical access as well as asset ownership, access to assistance, and market prices which were deemed relevant for economic access to food. The section on food security gave insight into the access-related causes of food shortage, and lastly, the section of remotely sensed geo-variables relates to food production and subsistence farming, a practice where farmers produce for their own consumption instead of selling to markets. For asset ownership, only assets related to transportation, cooking, and food storage were used. Sections on housing, education, health, employment, livestock, and non-farm enterprises, among others, were excluded from the analysis. Furthermore, indicator variables in each section were also filtered based on their relevance to food access as seen in Table 1.

Apart from HH survey responses, the geolocations of all 535 communities were also recorded. Geolocations of households, however, were not provided as these were considered sensitive and private information. Moreover, the process used by the DHS to further ensure privacy was adopted (Burgert et al., 2013) where a random offset was applied to the community locations which ranged between 0-2 km for urban communities and 0-10 km for rural communities. To maintain the accuracy of the locations, the displacement was constrained to remain within the same second administrative level (zones). The methodology section further describes how the random offset was addressed to ensure accuracy of the spatial representation even with the displacement.

Table 1. Ethiopia 2018 LSMS themes and variables used in study.

Theme	Collection Level	Variable name	Variable description
Identification	Community	ea_id	Enumeration Area ID
Access to basic services	Community	road_type	Type of main access road (1-4, best to worst)
		dist_aspRoad dist_bus dist_MUC dist_markets	Distance to (in meters): • Nearest asphalt road • Nearest bus station • Nearest major urban center • Large weekly market
		fare_MUC fare_woreda	Fare to (in Ethiopian Birr): • nearest major urban center • nearest woreda
Asset ownership	Household	bike_perc motor_perc cart_h_perc cart_a_perc car_perc	Do you own a: • Bicycle • Motorcycle • Cart – hand pushed • Cart – animal drawn

		ker_perc cyl_perc ene_perc refr_perc shelf_perc bio_perc ele_st_perc ele_mit_perc	• Private car • Kerosene stove • Cylinder stove • Energy-saving stove • Refrigerator • Shelf for food storage • Biogas stove • Electric stove • Electric mitad[2]
	Individual	phone_perc	• Mobile phone
Food security	Household	FI_expense_perc FI_transpoCost_perc FI_dist_perc	Causes of food insecurity: • Food in market was very expensive • Not able to reach market (high transportation cost) • Market very far
Access to assistance	Household	total_assist	Total value of: • Cash received • Food received • In-kind assistance received
Market prices	Community	price_1 price_2 price_3 price_4 price_5 price_6	Price of crop group in nearest market • Cereals • Pulses • Oil seed • Vegetables • Fruits • Tubers
Remote sensing geovariables	Community	twi sq1 sq2 sq3 sq4 sq5 sq6 sq7 af_bio_1 af_bio_12 cropshare popdensity anntot_avg ndvi_avg ndvi_max	Topographic Wetness Index Soil quality: • Nutrient availability • Nutrient retention capacity • Rooting conditions • Oxygen availability to roots • Excess salts • Toxicity • Workability Annual mean temperature (in °C) Annual precipitation (in mm) 2018 percent cropland in local area 2018 population density per km^2 Average annual total rainfall (in mm) Long-term average NDVI in primary growing season Long-term maximum decadal NDVI value in primary growing season

[2] Mitad is a flat pan primarily to prepare injera, a common staple food in Ethiopia.

2.3. Geospatial datasets

The following geospatial datasets were all clipped to the country boundary of Ethiopia and reprojected to the same coordinate system: WGS 84 Web Mercator, one of the most commonly used projection systems as well as the de-facto standard for web mapping applications (Battersby et al., 2014). This coordinate system was also used to ensure compatibility with outputs from different countries for possible future work. Since the main concern was on areas where people live, subsets of all datasets were made by removing areas with no population. The summary of geospatial datasets are found in Table 2 and the following sub-sections describe the datasets in detail.

Table 2. Geospatial datasets used in study.

Dataset	Source	# of variables	Use in study
Market locations	World Food Programme (2021a)	1	Create travel time to markets
Friction surface	Weiss et al. (2020)	1	Create travel time to markets
Travel time to markets	Computed in this study	1	Predictor variable for FAI
Road proximity	OpenStreetMap	1	
Bioclimatic variables	WorldClim 2 (Fick and Hijmans, 2017)	19	
Monthly precipitation	WorldClim 2 (Fick and Hijmans, 2017)	1	Predictor variable for cereal price
Travel time to cities	Nelson et al. (2019)	3	
Population density	WorldPop et al. (2018)	1	
Food prices	World Food Programme (2021b)	1	Predictor variables for FAI
Elevation	FABDEM (Hawker et al., 2022)	1	
Slope	Calculated from elevation	1	
Land use land cover	ESA (Zanaga et al., 2021)	1	
Soil quality	FAO and IIASA (2021)	14	
Agricultural production	MapSPAM (International Food Policy Research Institute, 2020)	23	
Urban-rural catchment areas	Cattaneo et al. (2021b)	1	Spatial aggregation of FAI

2.3.1. Market locations

Market locations in Ethiopia (119 markets) were provided by the World Food Programme (WFP) (World Food Programme, 2021a). Although this dataset does not contain all existing markets in Ethiopia, the markets in the dataset are those considered important by World Food Programme's (WFP) local offices based on regional representativeness and importance in WFP operations. Moreover, integrating crop price data to their respective markets was convenient because WFP's food price data also had corresponding market names which matched with the market names in the market location data.

2.3.2. Friction surface

As physical accessibility, in general, is largely dependent on the transport infrastructure, the updated friction surface (Weiss et al., 2020) provides some insight of the different aspects of such infrastructure like quality of roads, steep slopes, and land cover. It contained information on how many minutes it would take to travel one meter in a certain pixel. Road data came from a combination of both OpenStreetMap (OSM) and Google roads wherein Google roads helped maintain the connectedness of the road network in areas where availability of OSM road data was limited. Also, minor roads like unpaved roads were included in the development of the dataset which added a large volume of information that allowed for a more precise depiction of the transport infrastructure. The updated friction surface of Ethiopia is provided at approximately 1 kilometer resolution.

2.3.3. Travel time to markets

Adopting the established method to estimate travel time using the updated friction surface (Weiss et al., 2020) from previous studies (Cattaneo et al., 2021b; Nelson et al., 2019; Weiss et al., 2020, 2018), travel time to market locations (World Food Programme, 2021a) was calculated. Using the R package "gdistance" (van Etten, 2017), a transition matrix considering eight directions was created from the friction surface and then geographically corrected. From this transition matrix, the least cost path to market locations was calculated and the accumulated cost for that pixel is considered the travel time (in minutes) needed to reach the nearest market.

2.3.4. Road proximity

Physical access to food is the spatial accessibility of food sources relative to the access of transportation services of food consumers. The proximity of people to roads generally shows their level of access to the land transport network. The Euclidean distance from OpenStreetMap (OSM) roads in Ethiopia was calculated to create a road proximity dataset. To match the other geospatial datasets, the resolution was set to 1 kilometer.

2.3.5. Food prices

The WFP's Vulnerability Analysis and Mapping (WFP VAM) global food price database (World Food Programme, 2021b) is one of the largest global food price datasets, larger than both FAO's Global Information and Early Warning System on Food and Agriculture (GIEWS) and USAID's Famine Early Warning Systems Network (FEWS NET) (Food Security Information Network, 2015). The WFP VAM prices dataset for Ethiopia contained monthly food price data for several different crops from the year 2000 to 2021 for 110 different markets across Ethiopia. Not all markets had a complete monthly account of food price for all crops for the time period but the dataset is spatially and temporally sufficient to develop insights and investigate spatiotemporal trends in food price. Moreover, both retail and wholesale food prices are provided in the dataset. Monthly consumer price index (CPI) data (FAO, 2021a) were also collected to ensure that the food prices are comparable across time. The CPI is the measure that follows the cost of purchasing a basket of goods to the average consumer that may or may not change across time. Changes in CPI are usually associated to changes with the cost of living (ILO et al., 2004).

Cereal prices, out of the many crop prices included in the WFP VAM price data, were used in this study since cereals make up 68.3% of Ethiopia's total agricultural production (Seyoum Taffesse et al., 2011). The most recent spatial dataset of cereal price was from Cedrez et al. (2020b) who predicted the spatial and temporal variation of cereal prices across sub-Saharan Africa. However, the spatial resolution (~9 km) of their results was too coarse to be used in this study. Further, the price and market data they used from FEWS NET, GIEWS, and the Eastern African Grain Council's Regional Agricultural Trade Intelligence Network (RATIN) was quite limited for Ethiopia compared to the WFP VAM price data that was available. Hence, this study adopted and slightly modified their methodology for cereal price prediction using the WFP market and price data. In summary, both an annual average and monthly cereal price dataset were created for all of Ethiopia at 1 km resolution. Details on our cereal price prediction method and resulting data are found in the Annex (Chapter 7.1).

2.3.6. Urban-Rural Catchment Areas (URCAs)

URCAs (Cattaneo et al., 2021b) were used to determine how the disaggregated food access index varies across this new representation of the urban-rural continuum. This global dataset is a 1 kilometer resolution classification of settlement areas of different population sizes and their surrounding catchments in terms of 1, 2, 3 or more hours travel time. The settlement areas are: large city (population > 1 million), intermediate city (1million > population < 0.25 million), small cities and towns (0.25 million > population < 0.02 million), dispersed towns, and hinterlands. Figure 4 shows the total population living in each URCA while Figure 5 shows the URCA map for Ethiopia.

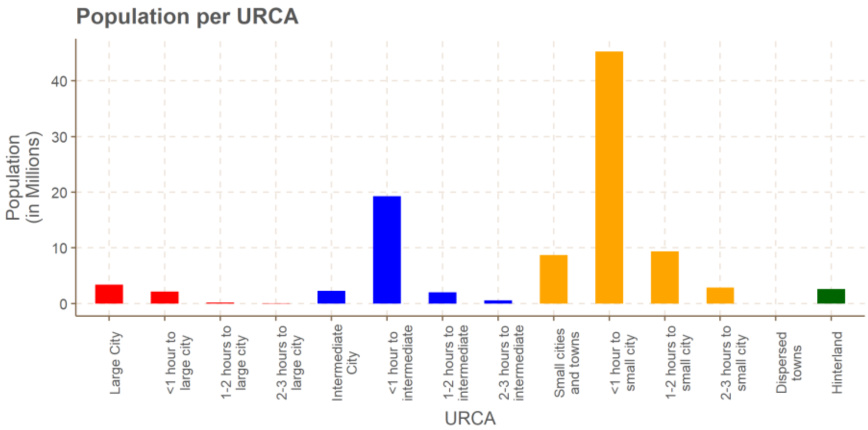


Figure 4. Population (in millions) per URCA.

This was used with the assumption that consumers in similar URCAs have similar food consumption behaviours and challenges. As this study focused on the consumer aspect of food access instead of the producer aspect, URCAs were deemed more applicable compared to other spatial strata like crop production system (CPS) zones which mostly concern the production aspect.

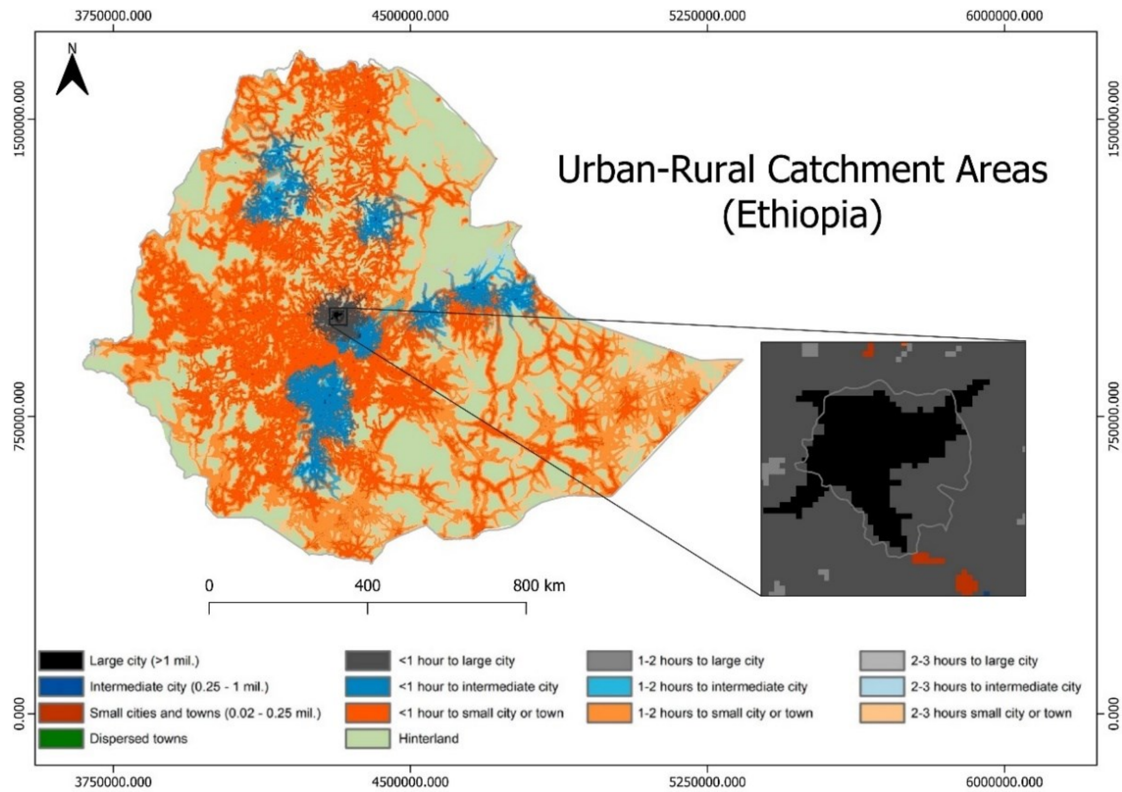


Figure 5. Urban-rural catchment areas in Ethiopia.

3. METHODOLOGY

3.1. Research methodology

The methodology was subdivided into three phases: construction of the food access index (FAI) from LSMS data, extrapolation of the FAI from geospatial covariates, and quantification of food access and food accessibility inequality across URCAs. Figure 6 displays the methodology flowchart.

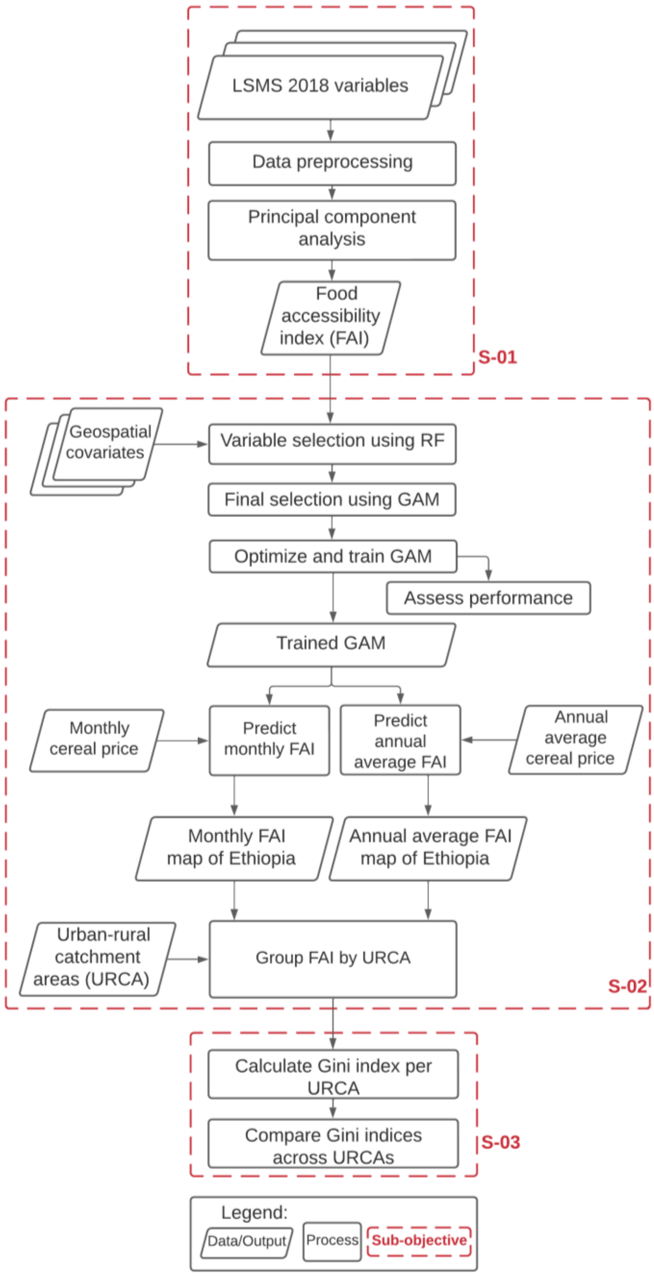


Figure 6. Methodology flow chart.

The subsequent sections provide more detailed descriptions of each step including assumptions and justifications.

3.2. Food accessibility index (FAI) construction from LSMS data

To construct the FAI, HH survey data pre-processing and principal component analysis were done as seen in Figure 7.

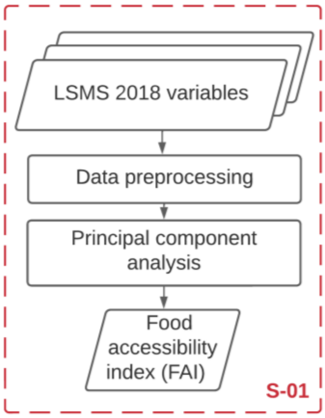


Figure 7. Method flow for FAI construction (Sub-objective 1).

3.2.1. Pre-processing of LSMS data

As responses in the LSMS data were collected at different scales (individual, household, community) but geolocations only at community level, there was a need to aggregate those responses collected at individual and household level in order to fit those responses to the geolocations. Aggregation was done by taking the percentage of responses per community so the results were comparable across communities. For example, for the indicator variable on ownership of a private car, the total number of households that own a private car in a certain community was divided by the total number of households in that community. In total, there were 485 LSMS communities with complete data. Some variables also needed further pre-processing which is further described in detail in the following sub-sections.

Access to assistance data was separated into three different types of assistance: cash, food, and in-kind. However, since not all communities had all types of assistance, there was a problem of incomplete data. To address this, the value of each type of assistance were summed up and this total was used in subsequent steps.

3.2.2. LSMS market crop price pre-processing

The market crop prices in the LSMS dataset were in the form of price per weight. However, some of the weights were expressed in local units (e.g. Birchiko, Joniya, etc.). To resolve this, LSMS also provided conversion factors to kilograms (kg) for each local unit and accounted for any variations across different regions. The weights were converted to kilograms based on these conversion factors and the price per kg was calculated.

Lastly, the calculated crop prices per kg were grouped into six crop groups: cereals, pulses, oil seeds, vegetables, fruits, and tubers. The average price per kg of all crops within each crop group in a community was calculated. Table 3 shows the grouping of crops.

Table 3. LSMS Market crop price groupings.

Crop	Crop Group
Wheat, Maize, Barley, Millet, Sorghum, Teff, Other cereal	Cereals
Field pea, Chick pea, Horsebean,	Pulses

Lentil, Haricot beans, Peanut/nut, Other pulse	
Niger seed, Linseed, Other oil seed	Oil seeds
Onion, Tomato, Kale/Cabbage/Lettuce, Other vegetables	Vegetables
Orange, Banana, Other fruit	Fruits
Potato, Sweet potato, Yam, Cassava, Godere	Tubers

3.2.3. Index construction using principal component analysis (PCA)

In order to construct a composite index to represent food accessibility, all the community-aggregated indicator variables were merged into one dataset. To construct the index, a weighted summation of all the variables was done where the weights of each indicator variable are determined using principal component analysis (PCA) (Hotelling, 1933) as proposed by Filmer and Pritchett (2001) and reinforced by its use in constructing the DHS wealth index (Rutstein and Johnson, 2004). PCA is a widely used multivariate technique for dimensionality reduction. It does so by converting, through orthogonal transformations, input variables into new variables that explain a unique portion of variance called principal components (PCs). Reducing the dimensions of our data in this way leads to faster performance with little to no cost to accuracy while also removing noise in the data. The orthogonal transformations also allow for the creation of uncorrelated features from the data. The R package "FactoMineR" v.1.34 (Lê et al., 2008) was used to implement the PCA while the package "ggplot2" v.3.3.5 (Wickham, 2016) was used to plot the scree and loading plots from the PCA results.

First, for all 485 samples, all 50 variables were standardized by calculating the z-score, which was done by subtracting the mean from each value and dividing by the standard deviation. This was done because PCA is sensitive to the range of values in the input data. Next, the covariance matrix of the dataset was calculated and the eigenvectors and eigenvalues were computed which were then used to determine the PCs. The orthogonal Varimax rotation was used to maximize the sum of the variance of the squared loadings or covariance. This was to ensure that each factor would only have a few high loading variables which would aid in interpreting what each factor represented. The first PC contains the maximum amount of information and the amount of information in the succeeding PCs decreases which allows for dimension reduction by discarding PCs that only contain minimal information or noise. The amount of variance in the dataset that each PC explains was visualized in a scree plot. Then, the coefficients of the variables contributing to each PC were analyzed.

One variable may have different weights or coefficients in each PC since a variable may have a significant contribution in one PC while a weak contribution in another. According to Hair et al. (1998), for a sample size of greater than 350, a coefficient or weight of greater than or equal to 0.3 is considered significant. This threshold was used to select significant variables in each PC and the next steps in the methodology proceeded only with these variables. Since each PC is a linear combination of the initial variables, the linear combinations were plotted to provide information on how the variables align on certain PCs. Specifically, the two PCs that explain the most variance in the dataset were visualized in a loading plot: the significant variables of the first PC on the x axis and the significant variables of the second PC on the y axis.

In the final step, only the weights from the first PC were used as this is assumed to be the best representation of the entire dataset among all other PCs (Filmer and Pritchett, 2001). Finally, following the calculation methods for the DHS relative wealth index (Rutstein and Johnson, 2004), the weights of significant variables from the first PC were multiplied by the standardized values of each variable and summed to calculate the FAI for each of the 485 LSMS communities.

3.3. Spatiotemporal food accessibility index prediction

The newly constructed FAI from Chapter 3.2 was integrated with the different geospatial datasets from Chapter 2.3. After which, based on the most important and significant geospatial variables, the FAI for the rest of Ethiopia was predicted over time. As the FAI only considers the consumer aspect of food access, all geospatial datasets were masked to populated areas, discarding uninhabited areas, similar to Chi et al. (2022). Figure 8 shows the detailed methodology flow to predict the FAI using geospatial data.

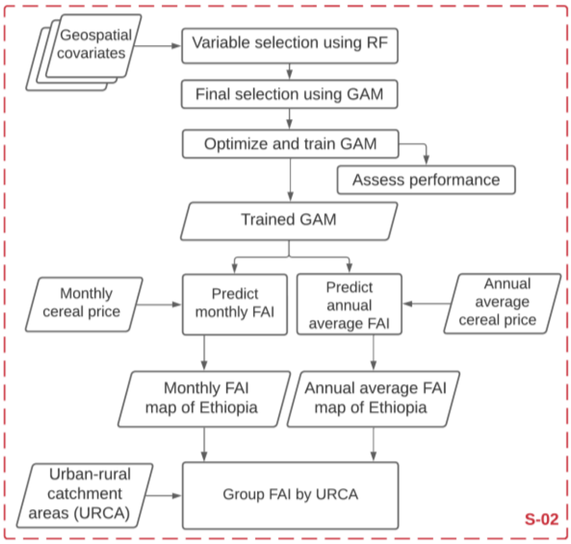


Figure 8. Method flow for FAI prediction (Sub-objective 2).

3.3.1. Spatial join and population weighted averaging

Each LSMS community survey record contained spatial markers (longitude and latitude coordinates). Using this, the newly calculated FAI values at each LSMS community were spatially linked to the digital geospatial data to be used in the subsequent processes. As random spatial offsets to the LSMS locations have been applied for privacy and ethical concerns (Burgert et al., 2013), there was a need to guarantee that the geospatial variables still accurately represented each LSMS geolocation. To address the 10 km maximum offset for rural areas and 5 km maximum offset for urban areas, Chi et al. (2022) used 4×4 grids for rural communities and 2×2 grids for urban communities since they used 2.4 km resolution data. This was slightly modified to 10×10 grids for rural communities and 5×5 grids for urban communities since our data was of 1 km resolution. Because all datasets were masked to cover only populated areas, there were some grids with incomplete cells, or cells with no population. Another modification to their method was that if the 10×10 or 5×5 grid had less than half of the cells it was supposed to, that LSMS geolocation would be discarded. For example, if a rural community only had 30 cells with a population of at least one, out of a possible 100 cells (10×10 grid), the community would be removed from the analysis. This is to ensure that the population weighted average is not biased and driven by only a few cells in the grid.

From these grids, the population-weighted average of all 62 geospatial predictor variables was taken and constituted the geospatial covariate data for each LSMS geolocation. After discarding 108 LSMS communities with a lack of cells with population in their grid, the final dataset of 377 samples with 62 geospatial predictor variables and the corresponding FAI as the response variable were created and used in the subsequent steps.

3.3.2. Variable selection using Random Forest

Random forest (RF) (Breiman, 2001) is a widely used machine learning algorithm that leverages several weak-learning regression trees. Each individual regression tree is considered "weak" since each tree only works with a random subset of samples as well as a random subset of features. These weak trees are then combined in an ensemble estimator using a bagging approach and the final prediction is based on a majority vote among the individual trees. In this study, RF was used in two separate parts of the methodology: predicting market cereal prices across Ethiopia and across months in the year (explained in the Annex, Chapter 7.1.2) and variable selection in the food access index extrapolation process.

The dataset created in Chapter 3.3.1 with 377 samples was split into training and validation datasets using a ratio of 70:30. The training set consisted of 265 samples and the validation set consisted of 112 samples. With the 265 samples for training and the large number (62) of predictor variables, variable selection was conducted to reduce the variables to only those that were important and significant. RF was preferred over other variable selection algorithms because it is known to be less prone to overfitting and was also found to outperform other machine learning algorithms in certain applications (Fernández-Delgado et al., 2014). Moreover, because of its performance and resulting features, RFs were even suggested to be the standard for gene selection (Díaz-Uriarte and Alvarez de Andrés, 2006) further reinforcing the reliability of RF for variable selection. In this study, in the context of interpretability, RF was also used to determine the importance of each variable when used in the model. This study used the R package "Variable Selection Using Random Forest" v.1.1.0 (VSURF) for the first variable selection (Genuer et al., 2010). The method of VSURF has three phases: thresholding, interpretation, and prediction.

- For the initial phase, also called thresholding, a minimal number of RFs was constructed and variable importance was calculated. Variables were ranked according to their mean variable importance, starting with those of highest importance. A decision tree was then fitted to the standard deviations of the variable importance values, from which a minimum threshold was calculated and those variables with importance below this threshold were discarded.
- During the second step, or interpretation phase, another set of RFs were built starting with only the variable with highest importance (from previous step) and ending with an RF where all the variables that were selected in the first phase are included, whereby producing the same number of RFs as the number of selected variables from the first phase. Another threshold was calculated using the minimum mean out-of-bag (OOB) error of the built RFs and their standard deviation. Of all the built models, the one with the least number of variables that had a mean OOB error less than "min error" was used and the variables of that model were selected in this phase.
- In the prediction phase, a stepwise selection of variables was conducted starting from the variables in the first phase. A variable was selected if its addition to the RF model resulted in a decrease in OOB error more than a threshold called "mean jump value". This threshold was calculated from the unselected variables in the interpretation phase and is the mean difference between OOB errors of a model and its succeeding model in a stepwise process.

Each phase had different considerations in selecting variables. For example, in the thresholding phase, only the variable importance was considered. The size of the model (number of variables included) was the main driver in the interpretation phase while the OOB error was used as a threshold. Lastly, in the prediction phase, the OOB error was optimized instead of being used as a minimum threshold. To check for substantial differences in performance, the OOB error and R2 were calculated for both a RF trained on all 62 variables and a RF trained only on the selected variables. The next step proceeded with the selected variables in the prediction phase.

VSURF focuses on the predictive performance of the model and not on the significance of each of the selected variables. As both performance and interpretability, including significance, are paramount in this study, an additional step to assess the significance of the importance metrics of the selected variables from the VSURF output was done using the R package "rfPermute" v.2.5 (Archer, 2021). The p-values of each variable's importance were computed by permuting the FAI (response variable) in a RF model. Variables were considered significant at $p \leq 0.05$ and insignificant variables were discarded from subsequent steps in the analysis.

3.3.3. Generalized additive models

The GAM was developed as a variant of the generalized linear model (GLM) where an additive function of smoothing operators is used instead of a simple linear function of the covariates (Hastie and Tibshirani, 1986). It is different from the GLM which uses parametric functions to fit data. The nonparametric nature of GAMs allows users to develop models without a priori knowledge of what types of predictor functions are needed. GAMs are flexible enough to handle both linear and nonlinear relationships between predictor variables and the response variable and also provides options to control the smoothing parameters for each predictor variable to avoid overfitting.

GAMs have been used for prediction in several applications because while it provides satisfactory predictions, it also offers clearer insights in terms of interpretability of results. Elith et al. (2006) predicted the spatial distribution of several species in portions of Australia, Canada, New Zealand, and South America using a GAM and concluded that GAMs outperform simpler methods. Crash frequency was modelled using a GAM and the relationship between crash frequency and road segment types were found to be both linear and nonlinear depending on segment type (Zhang et al., 2012). GAMs also outperformed GLMs in predicting potential landslide occurrences within the Sacheoncheon area in Korea (Park and Chi, 2008).

In the development of the model to predict the FAI, both RF and GAM were used. RF was used for variable selection because it is known to be an efficient and robust nonparametric algorithm that can deal with a high number of correlated predictors, nonlinear relationships, and interactions (Genuer et al., 2010). However, RF or any decision-tree based algorithm is not recommended for extrapolation, or predictions made outside the bounds of the training data (Hengl et al., 2018) hence the use of GAM which deals better with extrapolation. This is due to its smoothing nature of the data which follows the trend when dealing with data outside the training data range. This is useful in this study because the range of FAI values in the training data were not expected to be the full range of values for the entire Ethiopia. This is because the LSMS geolocated points, though nationally representative and well distributed across rural and urban areas, were enumeration areas, usually located near densely populated centers, hence missing out on the extremely remote areas with less population. The prediction model had to be dynamic and robust enough to deal with extreme values outside the training data range. Another reason why GAM was preferred over RF was that the relationships created in a RF may become very complex to interpret (Marshall et al., 2017).

In this study, a GAM was optimized and trained using the selected variables in Chapter 3.3.2 to (1) predict the FAI at 1 km resolution for all of Ethiopia and (2) demonstrate the functional relationship between each predictor and the FAI. The performance of the model was assessed using several accuracy metrics as well as the practicality of the results in terms of predictor-response relationship. The R package "mgcv" v.1.8-35 (Wood, 2011) was used to implement the GAM. The "'mgcv" package was preferred over the "gam" package (Hastie, 2020) because latter's method of estimating the smoothing parameters, the local scoring algorithm, is more computationally expensive than that of "mgcv" which uses an automated selection called penalized iterative re-weighted least squares (PIRLS).

3.3.4. Optimizing GAM input smoothing parameters

The 377 samples with only the 8 selected variables were first split into training and validation sets with a 70:30 ratio, resulting in 265 training samples and 112 validation samples. Since cereal price was one of the final variables selected, the annual average of cereal prices were used as input to the GAM. The inputs to the GAM were smoothing functions for each of the selected variables.

To optimize the smoothing function inputs to the GAM, three types of smoothing bases for each of the four smoothing functions were tested: thin-plate regression splines (Wood, 2003), cubic regression splines (Wood, 2017), P-splines (Eilers and Marx, 1996), resulting in 256 possible combinations of smoothing bases. The smoothing parameters, denoted by λ, for each smoothing function are internally optimized based on the data using a mixed model approach via restricted maximum likelihood (REML) because the commonly used generalized cross validation criteria (GCV) tends to under-smooth (Wood, 2017, 2011), and therefore overfit. To select the best model configuration, the Akaike information criterion (AIC) score was calculated for each combination as well as the significance of each smoothing function. The resulting trained GAM with the lowest AIC where all variables are significant at $p \leq 0.05$ was considered as the best performing GAM.

3.3.5. GAM assessment

To assess the interpretability and performance of the final GAM, three steps were conducted: partial dependence plotting, error metrics calculation, and comparison with a simple linear regression. The previously created validation set of 112 samples in Chapter 3.3.2 were used to assess the performance of the final GAM.

- First, the partial dependence of each variable from the final GAM was plotted to see the relationship between each predictor variable and response variable across the predictor variable's data range. These were first analyzed to see the practicality of the response curves.
- Next, the coefficient of determination (R^2) and root mean square error (RMSE), normalized RMSE (nRMSE), mean absolute error (MAE), and normalized MAE (nMAE) were calculated from the predictions of the final GAM and the validation set. The observed and predicted values were plotted to have a better understanding of how the final model performs on unseen data. The squared errors were plotted to see at which FAI values the model makes larger errors while the absolute errors were plotted to determine whether these errors were under- or overpredictions.

3.3.6. FAI spatial prediction

The annual average FAI was predicted for all of Ethiopia using the trained GAM. The FAI was then grouped by URCA and boxplots were used to visually assess the differences in FAI per URCA. Further, the total population of areas with FAI with (1) less than -10, (2) between -10 and -5, (3) between -5 and 0, (4) between 0 and 5, (5) between 5 and 10, and (6) greater than 10, was determined. The boxplots were created using the R package "ggplot2" v.3.3.5 (Wickham, 2016).

The interpretation of the FAI would be that higher index values meant better food access while lower values meant worse food access. The mean FAI for the entire Ethiopia was calculated to define "average food access". Though in reality, it is difficult to define as this can be subjective but in this study, the average of the four variables were calculated for FAI values equal to the mean $\pm$ 0.1 to be able to describe what "average food access" is, in terms of the four variables.

From the predicted FAI, a one-way analysis of variance (ANOVA) was done to determine whether there was a significant difference in the average predicted FAI across the different URCAs. After which, Tukey's range test was conducted to find out which pairs of URCAs had such significant differences.

3.3.7. FAI temporal prediction

The temporal prediction of FAI was done to assess how food access varies within a year assuming no major events like drought or conflict were to occur. Cereal price, as described in Chapter 2.3.5 and calculated in the Annex (Chapter 7.1.2), was the only temporally varying predictor from the selected variables. Thus, to predict the FAI across months in a year, the annual average cereal price was replaced with monthly cereal price for each of the 12 months while keeping the other non-temporally varying predictors as is. The FAI for each month was then predicted using the same trained GAM in Chapter 3.3.4. The pixel-wise absolute difference between each monthly FAI and the annual average FAI also calculated afterwards.

Similarly to the spatial prediction, ANOVA was conducted for each URCA to determine whether there was a significant difference in the average FAI across different months in each URCA. If there was, Tukey's range test was done to determine at which months had significant differences in average FAI for each URCA.

3.4. Food accessibility inequality quantification

From the predicted FAI for all of Ethiopia, food access inequality was calculated for each URCA. Figure 7 shows the methodology for calculating such inequality.

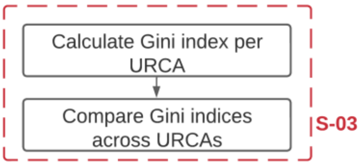


Figure 9. Method flow for food access inequality quantification (Sub-objective 3).

3.4.1. Inequality

Inequality can be interpreted in several different ways. A simple definition would be the opposite of equality, but there is too much nuance as to what inequality really is. Rein and Miller (1974) mentioned nine different standards of equality, two of which, this study relates to: complete horizontal equity and the goal of guaranteeing that nobody falls below a certain standard of living.

The first definition is relevant in terms of food accessibility because complete horizontal equity would be the ultimate goal in terms of equality, where everyone in the world has the same level of food access. In the context of income or wealth, in order to achieve perfect equity, a transfer of assets must be done from those that have more than enough to those that don't. However, this is not applicable to food access, since reducing the level of food access of those that have more than sufficient food access does not necessarily help those who have less food access. It is, however, possible to purely increase food access by improving the transport infrastructure, for example, which increases physical access. Yet, the context that inequality has been used in most studies view the asset as a diminishing commodity like income or wealth. And since food access is not a diminishing commodity, the way food access inequality is represented in this definition is not necessarily the best. Nevertheless, quantifying the level of inequality still aids in identifying problems in terms of food access like inequality in infrastructure development resulting in low levels of physical access or simply financial inequality leading to inadequate economic access.

The second definition is less about actual equality but more about a social minimum, which is a more realistic goal in terms of food access where everyone in the world has sufficient food access albeit unequal. This focuses on ensuring a certain level of food access which is sufficient for everyone. However, the definition of "sufficient food access" can be very nuanced as it may differ across many factors, be it location, culture, or even time (what was sufficient food access 10 years ago may no longer be true now). Nevertheless, this second definition does not represent food access as a diminishing limited asset that is unequally distributed. Food access is something that can be increased in a certain population group without the need to transfer from another.

In this section on inequality, this study focuses on the first definition. One way to measure the concentration of wealth was using the Lorenz curve (Lorenz, 1905), a visual representation of the cumulative normalized value of a certain asset, usually income or wealth, and the cumulative rank of the population that holds the least of such asset up to the most. It displays a diagonal line, sometimes called the line of equality which represents absolute equality, and the further the Lorenz curve is from this line, the greater the inequality. Also, it displays how much percentage of the overall assets is held by a certain percentage of the population.

3.4.2. Gini index

The Gini index (Gini, 1936), the degree of inequality in distribution, is derived from the Lorenz Curve. It generally measures how far the Lorenz curve is from the line of equality (Cowell, 2009). Based on Figure 10, the Gini index is the proportion of the area above the Lorenz curve, denoted as *A*, relative to the entire area under the line of equality. The range of values is 0 to 1 where 0 means the Lorenz curve is equal to the line of equality and values closer to 1 mean higher inequality. Although the Gini index was first defined to quantify income or wealth inequality, it has also been used to assess inequality in several applications like the use of water (Seekell et al., 2011) and energy (Lawrence et al., 2013). Bell et al. (2021) analyzed inequality in food system indicators across different countries using the Gini index and the degree of convergence was visualized using Lorenz curves.

The following equation is used to calculate the Gini index (G) from the Lorenz curve:

$$G = A/(A + B)$$

where *A* is the area above the Lorenz curve and *B* is the area below the curve.

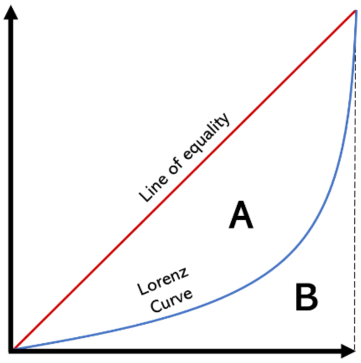


Figure 10. Diagram of a sample Lorenz curve.

While the inequality measured using these methods usually uses asset or income data from individuals, this study uses the FAI which represents communities (one pixel is one community). Therefore, the Gini index calculated in this study only represents communities and not individuals. The predicted FAI from Chapter 3.3 ranges from negative to positive values, but the Lorenz curve computes the cumulative percentage and therefore, requires non-negative numbers as input. To address this, the FAI was normalized to values between 0 and 1 by subtracting each FAI value with the minimum FAI value and dividing by the difference of the maximum and minimum FAI values.

In this study, the Gini index was used to quantify food access inequality in the communities of Ethiopia. The Gini index for the entire Ethiopia which represented the overall food access inequality was calculated using the normalized FAI. Food access inequalities across the different URCAs were also quantified by grouping the normalized FAI per URCA. This was then plotted to compare the food access inequality across URCAs. The R package “DescTools” v.0.99.44 (Signorell, 2021) was used to calculate the Gini index.

4. RESULTS

4.1. Food accessibility index (FAI) construction

4.1.1. Variable loading, alignment, and selection

In order to construct the FAI, there were 486 samples after aggregation with complete data for all 50 variables from the LSMS dataset. As seen in Figure 11, the principal components (PCs) created by the PCA showed that the first PC explained 16% of the variance in the dataset while the second, third, and fourth PC explained 10%, 6%, and 4%, respectively. The rest of the PCs were considered noise as these only explained around 3% or less.

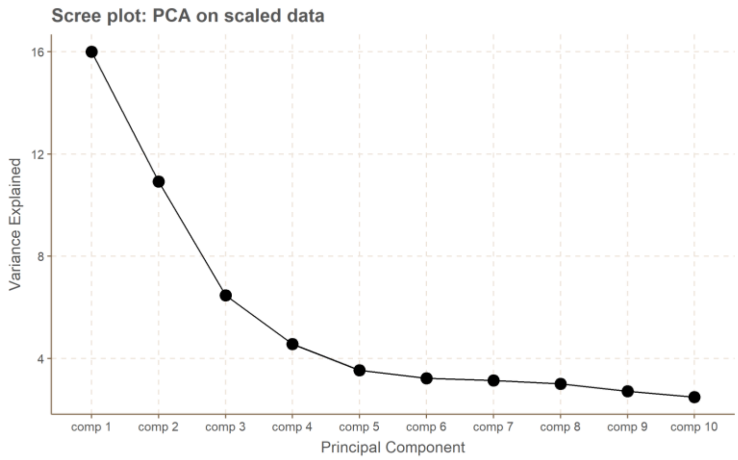


Figure 11. Scree plot of PCA showing first 10 PCs.

For the first PC, only 24 out of the 50 variables were significant: Ten of which were on asset ownership, seven were related to proximity and fare to services, three on soil quality, three on population, and one bioclimatic variable. The second PC had 15 significant variables. Ten of the significant variables were related to food production inputs, of which, 7 were on soil quality and 3 were related to temperature and precipitation. Three were related to food production outputs like the normalized difference vegetation index (NDVI) and the percentage of land used for agriculture. There was one variable on food price and one variable on asset ownership. Table 4 summarizes the coefficients of all significant variables in the first PC.

Table 4. Significant variables of first principal component.

Theme	Variable	Coefficient
Asset ownership	Electric stove	0.84
	Electric mitad[3]	0.83
	Refrigerator	0.82
	Phone	0.79
	Shelf for food storage	0.76
	Car	0.57
	Energy-saving stove	0.52
	Bike	0.42
	Kerosene stove	0.42
	Cylinder stove	0.31
Road quality	Road_type	0.65

[3] Mitad is a flat pan primarily to prepare injera, a common staple food in Ethiopia.

Population	Population	0.66
	Population density	0.65
	Number of households	0.56
Distance to services	Major urban centre	-0.51
	Asphalt road	-0.44
	Large weekly market	-0.36
	Bus station	-0.33
Fare to services	Woreda	-0.50
	Major urban centre	-0.42
Soil quality	Toxicity	-0.32
	Excess salts	-0.32
	Nutrient availability	-0.31
Bioclimatic	Annual mean temperature	-0.32

In the loading plot of the first two PCs in Figure 12, the variables on the right side of the plot were related to asset ownership, population, and the quality of the road. Those on the bottom side were variables on precipitation, normalized difference vegetation index (NDVI), and percentage of land used for agriculture. On the left side of the plot were those related to proximity and fare to certain services and another cluster of variables on the upper-left side of the plot were related to soil quality and temperature with one variable on asset ownership. Lastly, the variable on food price solely aligned on the upper-right side of the plot. Variable names can be found in Table 1.

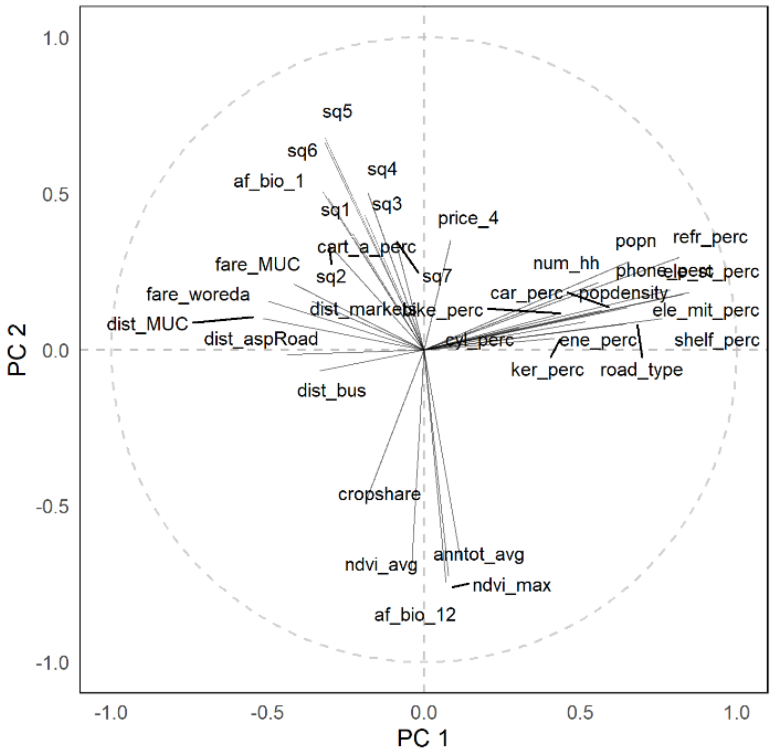


Figure 12. Loading plot of first and second PC.

Along the first PC, or x-axis, the variables with strongest positive coefficients observed were those related to owning electrical appliances like the electric stove and refrigerator where coefficients were above 0.8. Variables related to proxies for urban-rural areas like road type and population density also had relatively strong positive coefficients. The variables on distance to the nearest major urban centre and the fare to the nearest woreda had the strongest negative coefficients while other variables on proximity and fare to other services also had negative coefficients.

4.1.2. FAI calculation

After conducting the PCA and the weighted sum to construct the FAI, the resulting FAI had a range of values from -16 to 24. The higher positive range was likely because out of the 24 significant variables, 14 had positive coefficients while only 10 had negative coefficients. Additionally, variables with positive coefficients had greater magnitudes, reaching up to 0.8 for some, while the magnitudes of the negative coefficients only reached about -0.5. Even with the higher maximum value, as seen in Figure 13, the histogram of the FAI is right skewed and showed that most of the communities in the LSMS dataset of Ethiopia had FAI values below zero. Further, it was observed that most lie around FAI values of about -5.

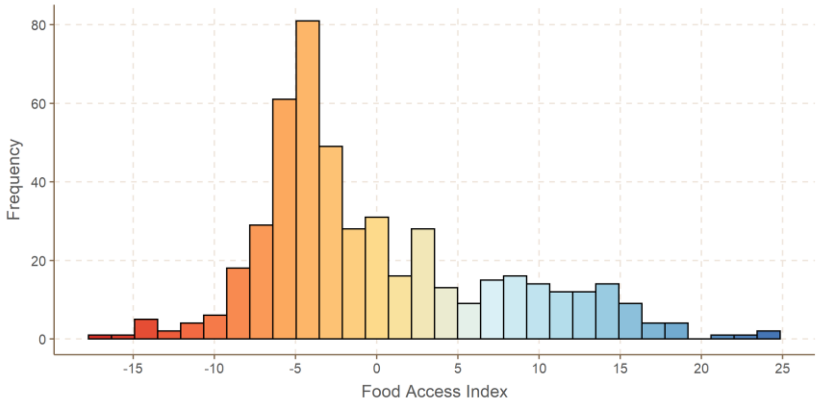


Figure 13. Histogram of constructed food access index.

The map of the FAI in each LSMS community in Figure 14 also shows that the highest levels of food access are concentrated in the communities located in the capital city of Addis Ababa and that the lowest levels of food access are found in the communities located in the south to southeast region of Ethiopia, in the Somali region.

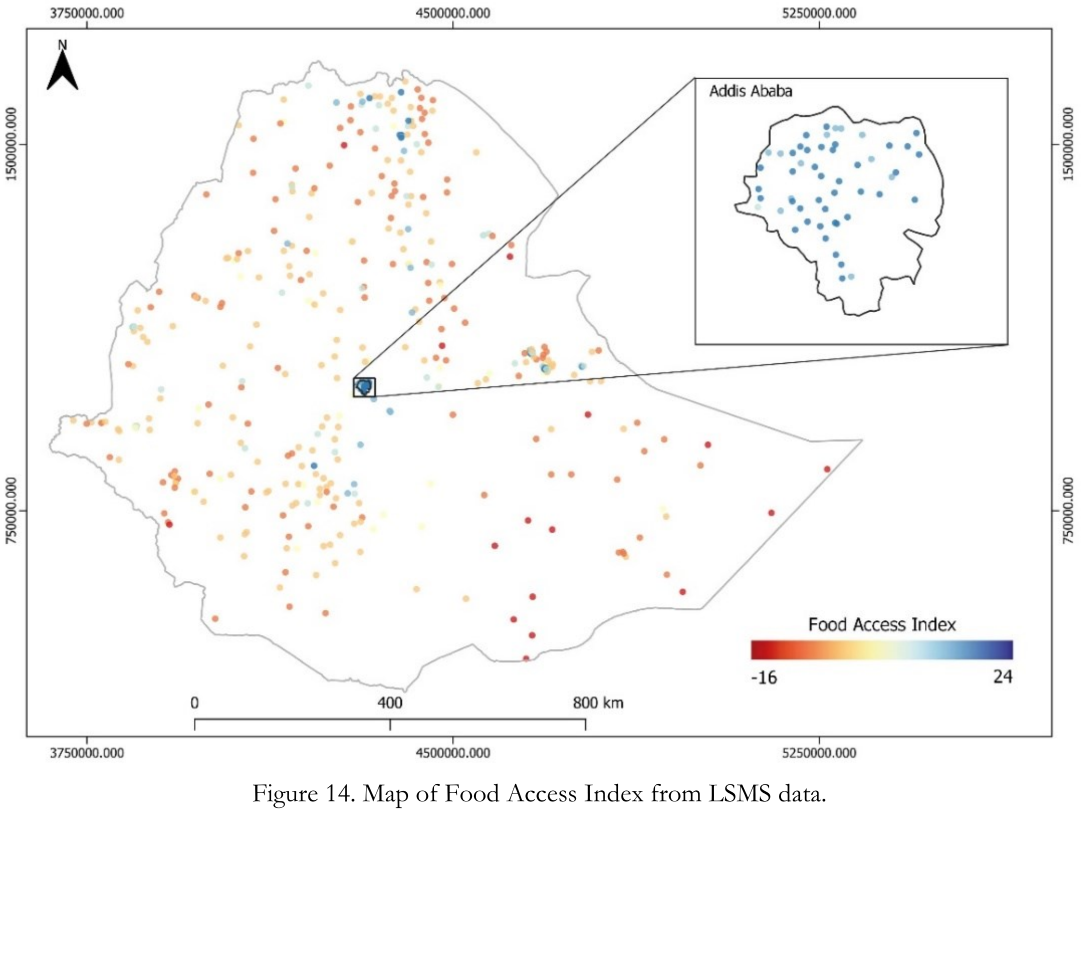


Figure 14. Map of Food Access Index from LSMS data.

4.2. Spatiotemporal food accessibility index prediction

4.2.1. Variable selection using RF

In VSURF, the first step of thresholding reduced the original set of 62 geospatial variables to 58 and during the interpretation step, from 58 to 21 variables. The least OOB error was achieved when using 21 variables as seen in the left plot of Figure 15. From the 21 variables in the interpretation step, only 11 were kept after the prediction step and upon calculating the significance of the importance metrics of the remaining variables using RF permutation, a final set of 8 variables were selected. The OOB error and R^2 of the initial RF with 62 variables was 12.7 and 0.77, respectively, while the RF with only the 8 selected variables attained 12.01 and 0.79, respectively. The right plot of Figure 15 shows the number of variables kept at each stage of VSURF and the RF permutation.

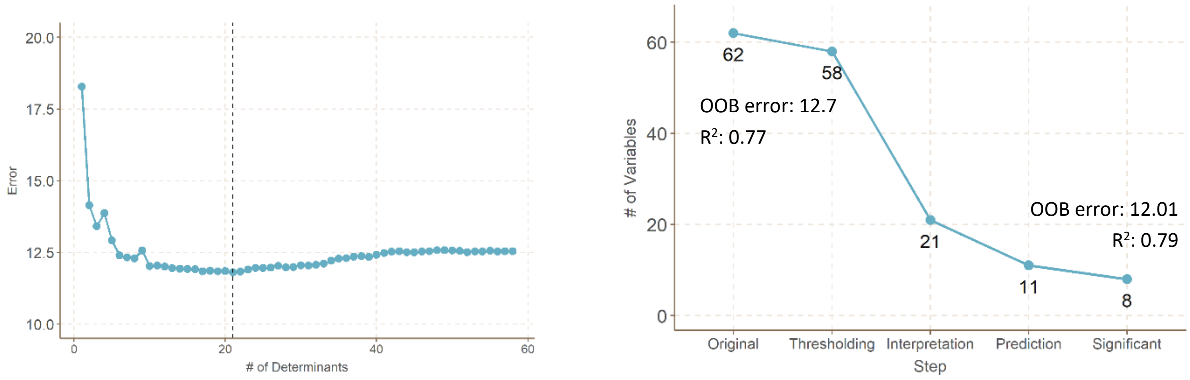


Figure 15. Error plot of interpretation step (left), summary of selected variables per selection step (right).

The 8 selected variables were road proximity (RPX), travel time to markets (TTM), slope (SLO), precipitation of coldest quarter (B19), cereal price (CRP), production of sorghum (SRG), beans (BNS), and barley (BRL). All 8 variables' importance metrics were significant at $p \leq 0.05$. Road proximity and travel time to markets had the highest importance while the other variables had relatively the same importance as seen in Figure 16.

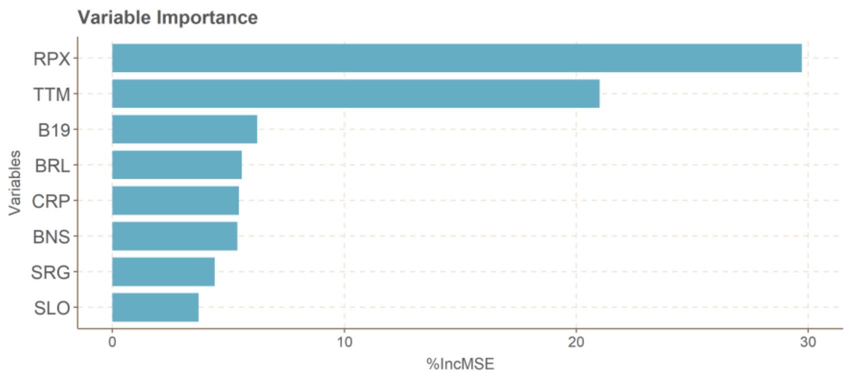


Figure 16. Importance of selected significant variables from VSURF and RF permutation.

4.2.2. GAM assessment

Upon training the GAM with the 8 selected variables from the variable selection step, it was observed that only 4 of the 8 variables were significant ($p \leq 0.05$) in the GAM. The variables BNS, SRG, BRL, and SLO were deemed insignificant in the GAM and were removed. In summary, from the 62 initial geospatial variables, only 4 variables were used in the final model after variable selection and GAM optimization. Only road proximity (RPX), travel time to markets (TTM), cereal price (CRP), and precipitation of the coldest quarter (B19) were significant in the final GAM and were used for training and prediction. After optimization, the best performing GAM used the cubic regression spline for RPX and TTM with smoothing parameters of 77.2 and 1, respectively, and used the P-splines for B19 and CRP with smoothing parameters of 142.5 and 79.9, respectively. In Figure 17, the partial dependence plots (PDP) of the FAI and each individual variable from the best performing GAM are shown.

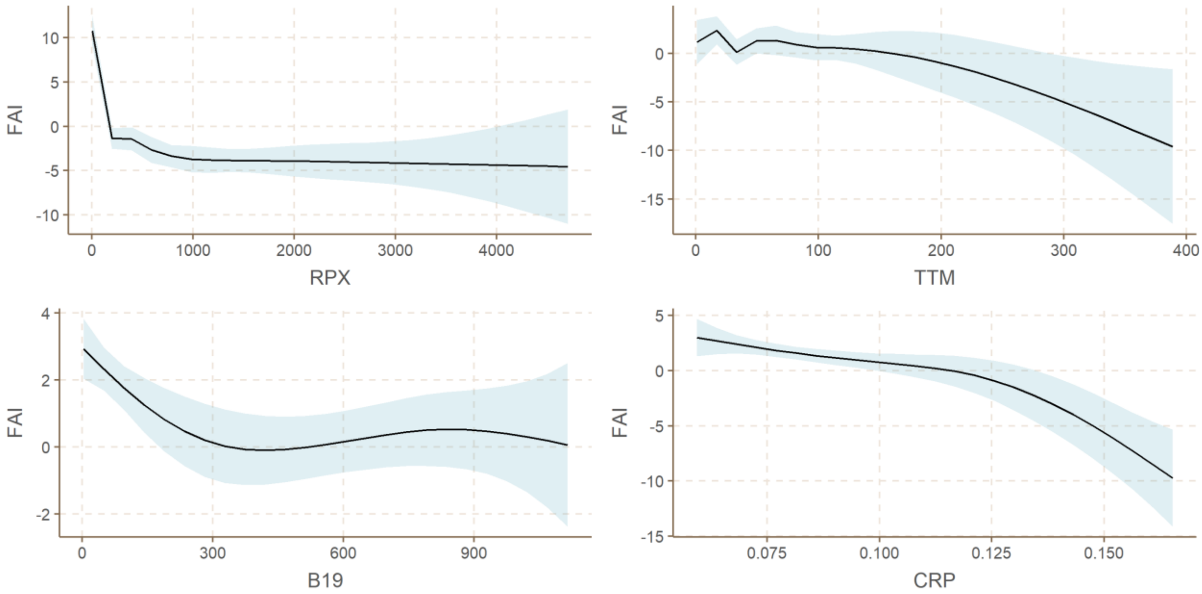


Figure 17. Partial dependence plots of final 4 variables used in GAM: road proximity (top-left), travel time to markets (top-right), precipitation of coldest quarter (bottom-left), cereal price (bottom-right).

For RPX, the FAI was initially high at zero distance from roads then steeply drops until about 200 meters then starts to flatten out from there but still slightly decreasing. In terms of TTM, FAI declines as travel time to the nearest market increases. The decline of FAI is steady until around 120 minutes travel time from which, the decline becomes steeper. B19 shows a similar behaviour with RPX but the initial decline is less steep. The FAI goes down until about 300 mm of precipitation and then slightly increases again up to 900 mm of precipitation but is relatively flat. Lastly, CRP shows a steady decline in FAI as price increases but this decline becomes more exaggerated after CRP reaches about 0.12.

Figure 18 shows three plots pertaining to the performance of the GAM: 1) observed-vs-predicted, 2) squared error, and 3) absolute error for the validation set to aid in understanding the performance of the GAM.

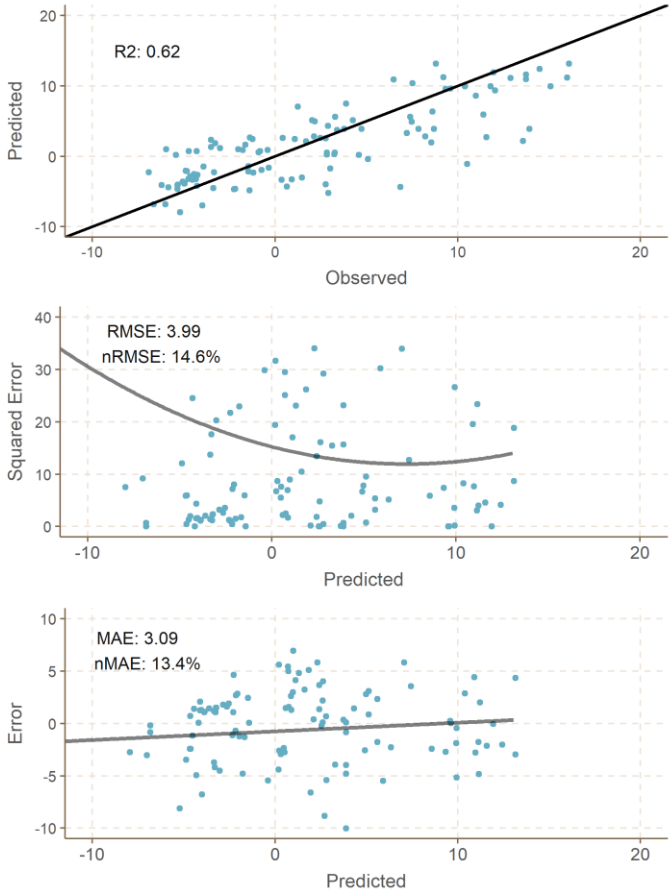


Figure 18. Performance plots of GAM for validation: observed-vs-predicted (top), squared error (middle), absolute error (bottom).

For the validation set, the observed-vs-predicted scatterplot shows a relatively tight fit suggesting relatively good performance by the GAM and further supported by its R^2 of 0.62. Looking at the squared residuals, the squared error plot shows the model error decreases as it predicts higher FAI values and the model attained a RMSE of 3.99 and nRMSE of 14.6%. Lastly, the absolute error plot for the validation set indicates the model generally underpredicts, albeit only slightly, and more so in low FAI values than in high FAI values. The final model achieved a MAE of 3.09 and nMAE of 13.4%.

4.2.3. FAI spatial prediction

The FAI for the entire Ethiopia was predicted using the trained GAM. Figure 19 shows the predicted FAI values across URCAs.

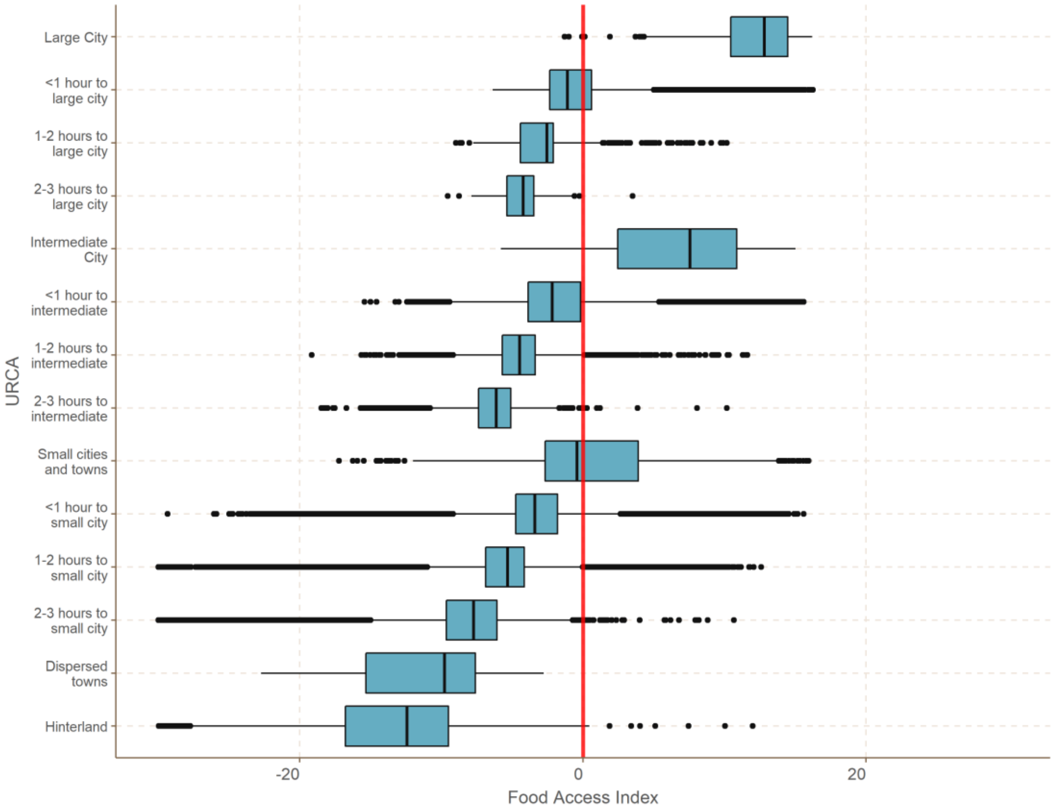


Figure 19. Boxplot of food access index per URCA.

As seen in Figure 19, the FAI levels in large cities and intermediate cities are well above zero which means their level of food access is relatively good. There are large differences in FAI levels from the urban settlements (large, intermediate, and small cities) and their respective immediate surroundings (<1 hour travel time) and dispersed towns and hinterlands have the lowest values of FAI. Also, there is a steep declining trend in FAI from the urban areas up to their catchment areas and this trend is more exaggerated in large cities and becomes less apparent as the population of the settlement area reduces.

The map of the FAI emphasized the spatial detail of the predicted FAI, as seen in Figure 20. The concentration of high FAI values are found near the capital of Addis Ababa in Ethiopia. Majority of the areas in Ethiopia have FAI values less than zero and the areas with the lowest FAI are those in extremely remote areas. Some areas directly surrounding the capital city, specifically the southwest, also have very low FAI values compared to its neighbouring urban areas as seen in the inset map of Figure 20.

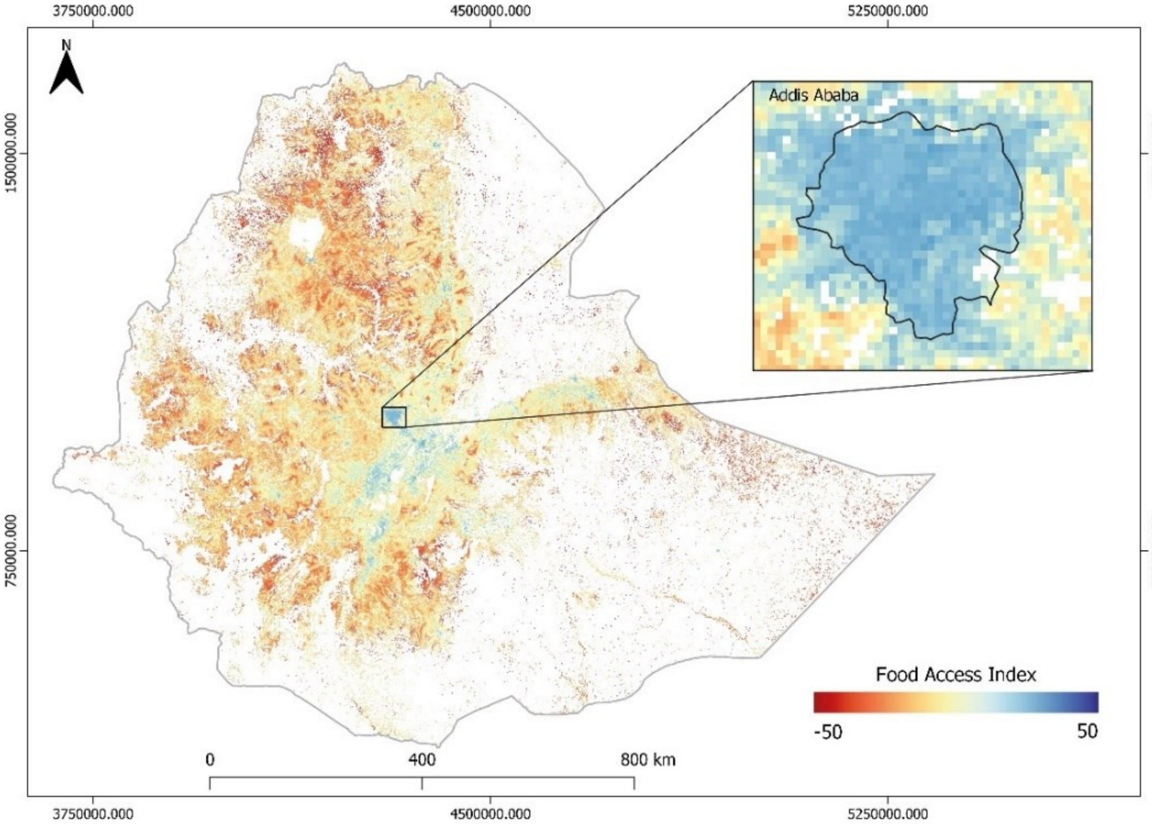


Figure 20. Food access index map of Ethiopia.

The impact of food access was quantified in terms of the number of people affected. Figure 21 shows that more than 50 million people have FAI values below zero. The mean FAI for the entire Ethiopia was -4.6. Average food access in Ethiopia, set at a FAI value ± 0.1 from the mean, as described by the average values of the four predictor variables, meant living about 1.7 kilometers away from a road, taking approximately 88 minutes to travel to the nearest market, experiencing a CPI-adjusted cereal price of 0.08, and having 435 mm of precipitation during the coldest quarter (December to February).

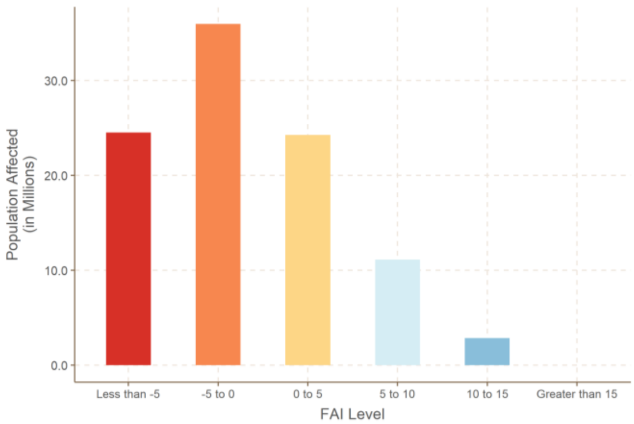


Figure 21. Population affected by each FAI level.

One-way ANOVA of the predicted FAI and URCAs showed that there is a significant difference in mean levels of FAI across URCAs at $p \leq 0.05$. The Tukey's range test also signified that all pairs of URCAs had significant differences at $p \leq 0.05$ except for two pairs:

- 1-2 hours to intermediate city and 2-3 hours to large city
- Dispersed towns and 2-3 hours to small city.

4.2.4. FAI temporal prediction

After predicting the annual average FAI for Ethiopia, the monthly FAI was also predicted using the respective monthly cereal prices instead of the annual average. The difference from the annual average FAI for each month was also computed. Negative differences meant the monthly FAI was less than the annual average, while the opposite for positive differences. The range of the difference across all months was from -7 to 3 but after the histogram of the differences across all months was inspected, as seen in Figure 22, majority of the values are from -2 to 2 so the monthly maps were visualized using this range of values.

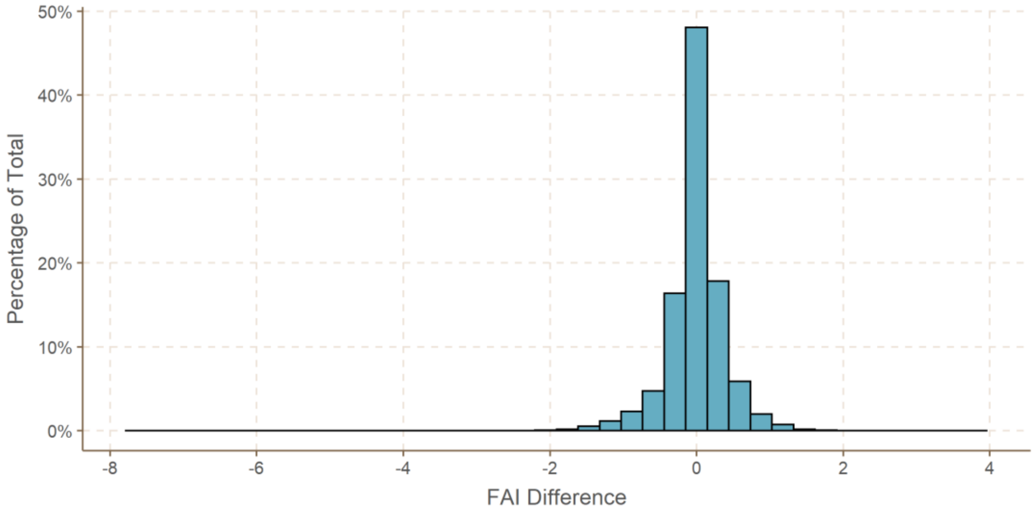


Figure 22. Histogram of FAI differences across months.

The monthly FAI predictions were also mapped. Figure 23 shows the monthly FAI difference across all of Ethiopia. In the month of January and February, the northwest region of Ethiopia has lower FAI values than the annual average while the central part has slightly higher FAI. In July and August, there is a large reduction in FAI in the central area of Ethiopia but also recovers by September until November where there is not much difference with the annual average. Lastly, in December, the FAI is at its highest in most parts of Ethiopia except for the far east and north west regions.

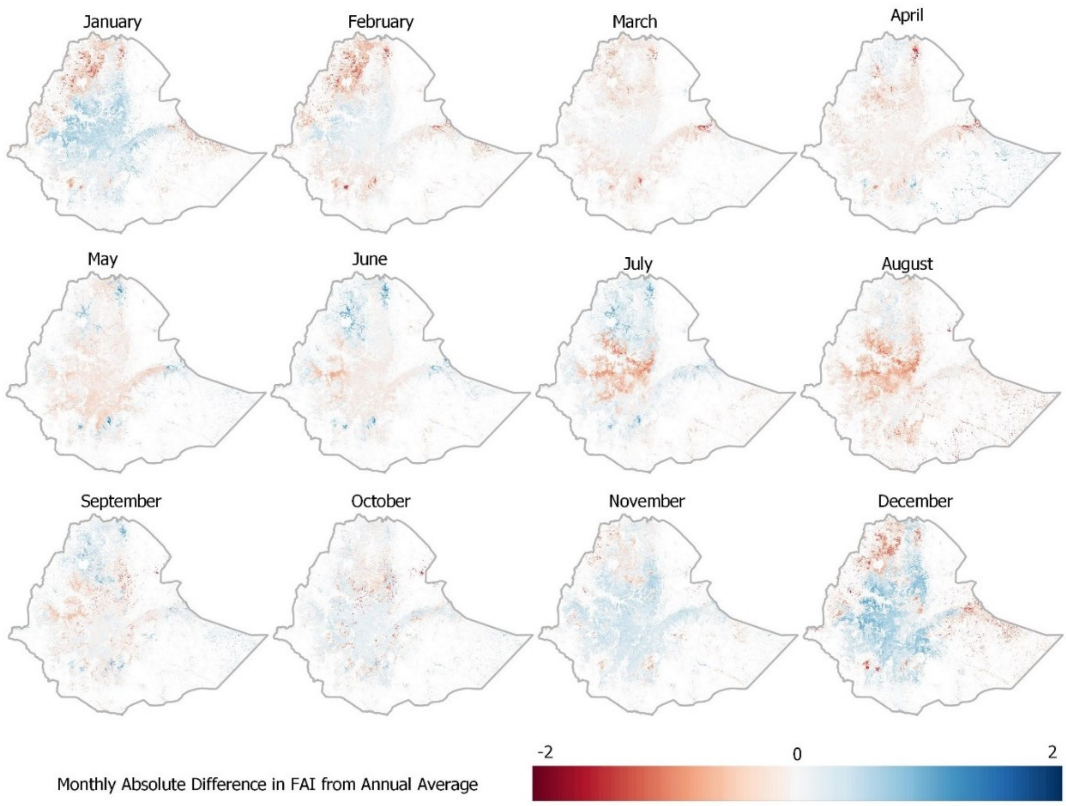


Figure 23. Maps of the absolute difference between monthly FAI and annual average FAI.

The results from the ANOVA and Tukey's range test conducted for each URCA showed that not all URCAs had significant differences in monthly average FAI from the annual average. For the URCAs that did, it was only true for some months in the year. In Table 5, a cell was colored red if the average monthly FAI for that URCA and that month was significantly ($p \leq 0.05$) less than the annual average and green if it was significantly more than the annual average.

Table 5. ANOVA and Tukey's range test results for monthly FAI per URCA.

URCA	**ANOVA ($p \leq 0.05$)**	**Tukey's ($p \leq 0.05$)**											
		J	F	M	A	M	J	J	A	S	O	N	D
Large city	Yes												
< 1 hr to large city	Yes												
1-2 hrs to large city	Yes												
2-3 hrs to large city	Yes												
Intermediate city	No												
< 1 hr to intermediate city	Yes												
1-2 hrs to intermediate city	Yes												
2-3 hrs to intermediate city	Yes												
Small cities and towns	Yes												
< 1 hr to small city	Yes												
1-2 hrs to small city	Yes												
2-3 hrs to small city	Yes												
Dispersed towns	No												
Hinterland	Yes												

There was no significant difference in the predicted FAI across all months in intermediate cities, 2-3 hours to intermediate city, and dispersed towns. Although the ANOVA result for 2-3 hours to intermediate city was significant, the pairwise Tukey's range test showed no significant differences between monthly average FAI and the annual average FAI. There were significant differences in monthly FAI across different months but this was not the focus of this study.

The predicted FAI for the month of August was found to have significantly lower monthly FAI than the annual average in 10 of the 14 URCAs, the most among all months. Additionally, our results from the Tukey's range test also indicated that the monthly FAI in peri-urban areas (< 1 hour travel) of intermediate and small cities were the most volatile as these had significant differences from the annual average for 10 out of the 12 months while catchment areas further away and the urban centers themselves had less temporally volatile FAI. Also, hinterlands had significantly lesser monthly FAI in December to February. In all periurban areas, October to January showed significantly higher FAI while the opposite was true for periurban areas of intermediate and small cities from February to April.

4.3. Food access inequality quantification

The Gini index was calculated to quantify food access inequality in a single metric. Figure 24 shows the calculated Gini index per URCA where catchment areas are coloured by their respective major settlement area: large city (red), intermediate city (orange), small city (blue). Dispersed towns and hinterlands were categorized together and coloured green.

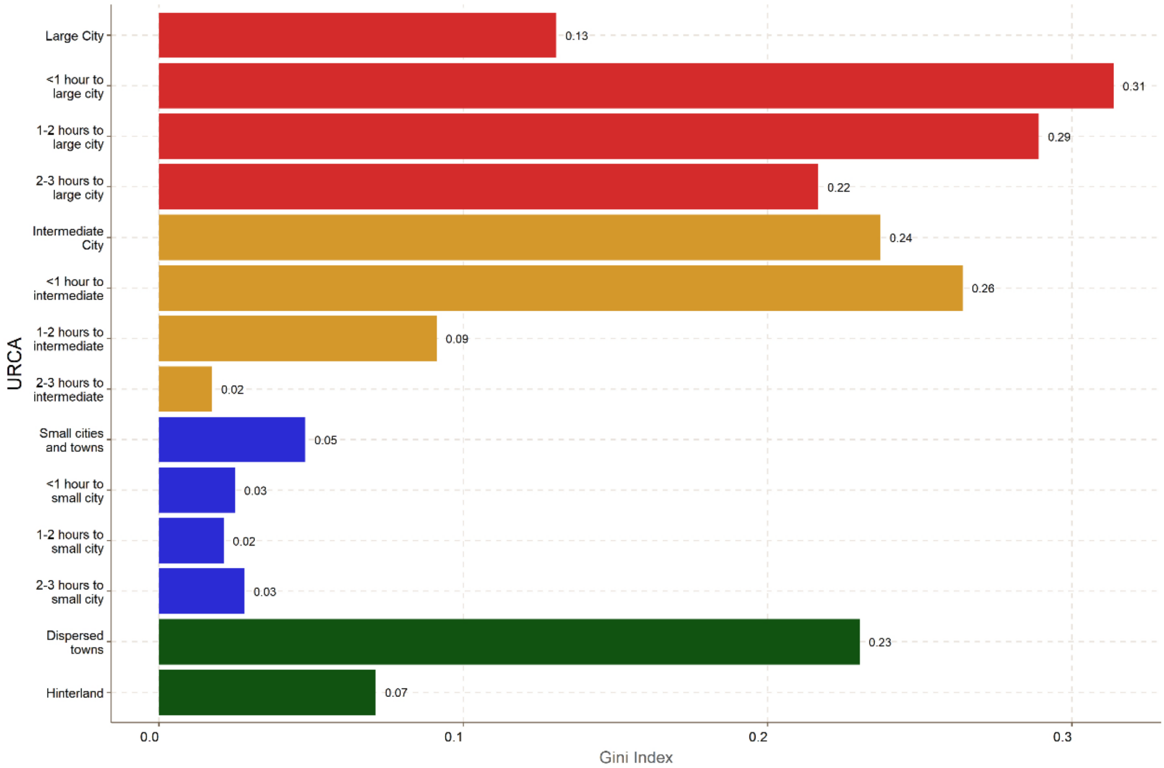


Figure 24. Gini index per URCA.

The calculated Gini index for the entire Ethiopia was 0.03 indicating that overall food access inequality in Ethiopia was low. However, calculating the URCA-specific Gini indices uncovered the variation in inequalities of food access. The peri-urban area (< 1 hour travel time) of the only large city in Ethiopia, the capital city of Addis Ababa, had the highest level of food access inequality with a Gini index of 0.31. The capital city itself, had a Gini index of 0.13, less than its catchment areas. Among the major settlement areas, dispersed towns and intermediate cities both had the highest levels of inequality. The major settlement areas with the least food access inequality were hinterlands and small cities with Gini index values of 0.07 and 0.05, respectively.

For large cities, the Gini index within the city was the least compared to its catchment areas but increased dramatically in its immediately surrounding area (< 1 hour to large city) and slightly reduced as time to travel to the large city increased. Further, food access inequality is high both within intermediate cities and its immediately surrounding areas (< 1 hour to intermediate city) and then lessens considerably in areas that take 1-3 hours to the nearest intermediate city. Meanwhile, small cities and their catchment areas had the least food access inequality with Gini index values ranging from 0.02 to 0.05.

5. DISCUSSION

5.1. Summary

Food accessibility is one of the four pillars of food security but has received less attention in academic literature compared to other pillars of food security like food availability. This thesis contributes to the food access literature by developing and demonstrating a method to estimate and map food access at high spatial detail across time for the whole country of Ethiopia. Two aspects of food access, physical and economic, were considered in estimating and mapping overall food access. To the best of our knowledge, this is a first in the current literature as food access has been spatially represented as only physical access (Chen, 2017; Jeong and Liu, 2020; Pozzi and Robinson, 2008) and both economic and physical access (Breyer and Voss-Andreae, 2013; Jiao et al., 2012), but only for one city or town. To do this, several methods were integrated to link ground-collected survey data of limited spatial scope and geospatial data which covers large areas. Doing this not only provides information on food access at a large spatial coverage, but also opened the opportunity to investigate the differences in food access across urban-rural catchment areas. The disparities in health and education across the urban-rural continuum have been examined (Cattaneo et al., 2022) but not of food systems. It is important to grasp the dynamics of food systems, specifically food access, from urban to rural areas as there are existing gaps on regional- and country-scale food access studies (Abu Hatab et al., 2019). However, locating the problem is only the first step towards a solution and a better sense of what type of problems are present in a certain location could be the logical next step. Apart from mapping food access, this research also sought to identify the inequalities in food access and what important factors contribute to food access. Such insights could be useful to evaluate what types of interventions can be made to improve food access in specific areas.

In this study, a new FAI was constructed from survey data where variables for both economic access and physical access were the strongest contributors. Of the two, economic access had a larger contribution which only further proves the need to look at food access from more perspectives apart from physical access, where most research on food access has been done. Food access, represented by the FAI, was found to be alarmingly low in peri-urban areas as seen in Figure 19, which constitutes about 67% of the population. Furthermore, food access was also more temporally volatile in these peri urban areas. Moreover, there was a clear disparity and trend in terms of food access levels between urban centers and their catchment areas. Similarly, differences in food access inequality were also emphasized, where inequality was high in catchment areas of large and intermediate cities but very low in small cities and their catchment areas.

5.2. Findings

The findings from the PCA of the LSMS data revealed that food access is mainly driven by both physical and economic access to food. Furthermore, economic access was a larger factor than physical access given the higher absolute loadings of variables relating to asset ownership in the first PC compared to the loadings of variables relating to physical access. The positive coefficients of the asset-related variables meant that when a community or household owned more electric appliances and assets and live in more developed areas, the food access index was higher. Meanwhile, the negative coefficients of the variables related to proximity and fare to certain areas signified that the further one was or the more one paid for fare to travel to such areas like a major urban center or an asphalt road, food access was lower. Also, variables related to asset ownership had stronger coefficients than those related to proximity to services. Thus, being located near roads or a major urban center did not necessarily mean high food access if asset ownership was low. This is an important insight as our findings showed that while spatially estimating physical access to food is important, economic access to food is equally, if not more, important.

Several geospatial variables were discarded in the variable selection process. All 14 variables on soil quality were not considered as important in quantifying food access as those on asset ownership and proximity to services. Agricultural production of 20 crops were also removed likely because (1) most of these were not major crops in Ethiopia and (2) the original spatial resolution of the crop production data was 10 km, relatively coarser than the rest of the geospatial datasets. From the 19 bioclimatic variables, only one was kept in the process. An overarching reason for the removal of these three families of variables (soil quality, agricultural production, bioclimatic) is that they were more related to food production and availability than food access. Nevertheless, these variables may still be important in quantifying food access but other collinear variables just explained more variance in the data. The eight significant variables selected by the RF were a mix of proxies for urban-rural areas, physical access, economic access, and environmental. First, road proximity and three agricultural production-related variables (barley, sorghum, and beans) were considered proxies for urban and rural areas as rural areas typically have higher agricultural production compared to urban areas. Furthermore, urban areas are expected to be nearer many roads compared to rural areas which have limited options. The GAM was able to further reduce the number of variables selected by RF from 8 to 4 with little to no trade-off in terms of model performance. RPX was retained over the roads were identified as key indicators for settlement agglomerations and areas closer to roads are also more favourable for social development (Liu et al., 2022). Also, slope was already a factor in the friction surface used to create the travel time to markets which could mean that travel time to market explained more variation than slope, hence the removal of slope.

In interpreting the PDPs of each predictor variable, apart from the cereal prices PDP, the confidence bands of the other variables were relatively wide nearing the extreme values which would be attributed to the lack of data points in areas that are far from roads and take longer time to travel to markets. This is expected as these are the same geolocated data points from LSMS used in constructing the FAI where the locations are enumeration areas and these generally have higher populations and are nearer roads. Nevertheless, the PDPs should be interpreted with caution especially in the limits of the training data range. The PDP for road proximity highlighted the importance of living very close to a road in terms of food access and how after a certain distance (~200 meters), food access is no longer substantially affected by increasing distance. It indicates that someone living very far from a road, 2 kilometers, for example, would not have a big difference in food access with someone living 3 kilometers away, considering the other three variables are kept constant. Meanwhile, there would be a substantial difference in food access with someone living in the city, where there are a lot of roads, and someone living in a peri-urban area where the nearest road could be a few hundred meters away. In terms of travel time to markets, the PDP revealed a different trend to that of road proximity. It emphasized the value of market access because as people take more time to get to markets, the downward and almost linear trend of the PDP after ~120 minutes suggests that food access only gets worse. The cereal prices PDP showed that while food access steadily declines as food prices increase, which is expected, it also delineated that at a CPI-deflated price of ~0.12, food access starts to plunge dramatically. It demarcated the tipping point of affordability of cereals and this is crucial since cereals are part of the daily diet of most Ethiopians and if cereal prices spike past this “limit”, a large part of the population may suddenly no longer afford it. However, as the cereal prices used in the model were an average of prices from 6 major cereal crops (see Table 3) and were also CPI-deflated, this price “limit” must only be interpreted as an indicative value and not an actual price value. Lastly, precipitation of coldest quarter is an environmental variable that affects food access where more precipitation during this period (December to February) meant less food access. The FAO crop calendar of Ethiopia (Figure 3) shows that this period is during or right after the harvest of most crops grown in the Meher season and when these crops are being transported to markets. The decrease in food access as precipitation during this time increases could be attributed to transport difficulties brought about by exacerbated poor road conditions when floods happen during the rainy season (Olana et al., 2018). This is also the period right before the short rains season, or Belg season, when secondary crops are grown. which could be speculated that if there was a lot of precipitation before the Belg season, the soils may be oversaturated during the sowing time of cereals. Although this does relate more to food production and less to food access, this could also be a proxy for subsistence farming where farming households produce their own food, therefore still having access to food without needing to travel to markets to buy their food.

The final trained GAM, using only the four selected variables, was able to explain 62% of the variation in the FAI, which is comparable to state-of-the-art methods in estimating wealth from geospatial data with similar frameworks, which explained between 56% to 75% of variations in what each was predicting (Chi et al., 2022; Jean et al., 2016; Yeh et al., 2020). Comparisons are made to methods estimating wealth as this study is the first to do so for food access. To further highlight the performance of the GAM in this study, these wealth prediction and mapping studies employed advanced and sophisticated methods like deep learning. These approaches generally perform well but require a large amount of training data and also provide difficulties in terms of interpreting what goes on within the model, although recent advancements have been able to give more insights in terms of model explainability (Barredo Arrieta et al., 2020; Molnar, 2019). On the other hand, the GAM used here, trained only on 265 samples and four variables, is a relatively simple model, easy to interpret, as seen with the partial dependence plots, and was still able to achieve comparable results. The squared error plot showed increasing error as predicted FAI values increased. From a practical perspective, making larger prediction errors in areas with generally high food access is not a critical issue because for planning interventions, areas with low food access are more important anyway, and in these areas, the GAM performs quite well. Despite the good performance, GAM does not account for non-linearities and extrapolates almost linearly at extreme values, which may result in extremely high or low predictions compared to the training data ranges. This explains the wider range of values of the FAI extrapolated by GAM compared to the constructed FAI from the PCA. Thus, care should be taken when interpreting these predictions since they are likely overexaggerated by the linear nature of the model's predictions at these extreme data ranges. Lastly, other simpler approaches like a multilinear regression model (MLR) could also have been used to predict the FAI. To justify the use of a more complex model like GAM, a MLR was trained on the same training set using the same selected variables. As seen in Table 7 in the Annex (Chapter 7.2), the GAM performed better than the MLR in terms of R^2, RMSE and nRMSE on the same validation set indicating that the relationships between the four selected variables and the FAI are indeed nonlinear and that predicting the food access index is a much more complex problem for a MLR to solve.

The predicted FAI of Ethiopia highlighted the large disparity in food access between major urban centers and their peripheral areas. Cattaneo et al. (2021b) emphasized the importance of peri urban areas (< 1 hour travel time to urban centre) of large, intermediate, and small cities because a substantial percentage of people live here. This is also true in Ethiopia where approximately 67% of the population live in peri urban areas of large (2%), intermediate (19%), and small cities (46%). This is also in agreement with Cattaneo et al. (2021b) where, in low income countries like Ethiopia, intermediate and small cities play a larger role in development than large cities because more people are located in the catchment areas of intermediate and small cities than in the catchment areas of large cities. Thus, based on our estimates, a sizeable part of the population have very low food access. Furthermore, in these same peri urban areas, the average travel time to the nearest market is about 1 hour: the same time it takes to travel to the nearest urban centre, which suggests that there may be limited options in terms of markets within the peri urban areas or the nearest market is in the urban center itself.

Economic access to food is naturally more volatile across time compared to physical access because transport infrastructure does not generally alter across months unless a major disaster or conflict limits access or a major investment in infrastructure improves it. With this, focus was placed on the temporal dynamics of economic access to food and overall food access. Predicting the monthly FAI of Ethiopia showed the impact of cereal prices towards food access. The differences in monthly cereal price and monthly FAI from their respective annual averages showed that cereal prices are highest and the FAI is lowest in the months of July and August. These months coincide with the middle of the "lean period" in Ethiopia's crop calendar (Figure 3), or the fragile period between planting and harvesting where most of the major crops are still growing. During this time, food stocks are limited and prices increase. This indirectly suggests that the FAI successfully captured this temporal phenomenon and given the good performance of the GAM to predict the FAI in space, the areas in central Ethiopia that show substantial reductions in the FAI in the months of July and August can be relied upon. Meanwhile, the hinterlands in the northwest region of Ethiopia suffered from higher prices and lower FAI from December to February while the central, more urbanized, areas of Ethiopia faced the complete opposite. This is also in line with Headey et al. (2019) who found that the rural poor population of Ethiopia live near markets that sell less diverse options at higher prices. At the same time, urban and peri urban areas of Ethiopia enjoy lower food prices because the markets they live near to are able to sell at lower prices.

To add to the already low FAI values in peri-urban areas, these areas also have the most temporally volatile FAI as seen in Table 5. Apart from August, previously mentioned to be the middle of the "lean period" of Ethiopia's crop calendar, food access in the peri urban areas of intermediate and small cities was also significantly lower than the annual average from February to April. The low FAI in these areas at this time of the year corresponds to the post-harvest time for most crops during the Meher season. This also coincides with findings from Hill and Fuje (2020) where they found considerable increases in grain prices from the end of January until April, mainly driven by rainfall shocks. Their findings, however, did not mention which specific areas in Ethiopia where affected by such price increases. Our results suggest that peri-urban areas of intermediate and small cities feel the impact of this phenomenon the most. The opposite was true from October to January in all peri urban areas, where food access was significantly higher than the annual average. This matches with the harvest season for most crops grown in the Meher season. In the harvest season, farming households generally subsist on their own production rather than purchase from markets (Hirvonen et al., 2016), reducing the demand. With this, food supply in the markets may be sufficient or even in surplus causing prices to be relatively lower at this time of the year thus people have higher economic access. Hill and Fuje (2020) also agree as they mention that the impact of rainfall shocks towards food prices gradually decrease from May to October and only start increasing from the end of January. These findings of the temporal volatility in food access show how food access in peri-urban areas is highly dependent on food prices leaving them vulnerable to price spikes which could be caused by a myriad of things like conflict, social uprising and drought. Abay and Hirvonen (2016) found that households located nearer to markets did not rely as much on producing their own food which can place some peri-urban areas in a challenging situation at certain times of year since they are not as close to markets compared to their urban counterparts and not all households produce their own food.

The overall Gini index, computed at community level and not at individual level as usually done, in terms of food access inequality for Ethiopia was very low. However, the levels of food access inequality were uncovered after computing the Gini index for each URCA. D'Odorico et al. (2019) stated that computing within-country inequalities in food access is important to tackle because even if one country has enough food to feed the whole population, this does not guarantee that the everyone does indeed have access, either physically or economically, to food because of the large disparities that exist. Our findings not only provide quantification of within-country inequalities in food access but also stratified by URCAs. Food access inequality in the catchment areas was positively correlated with the size of their respective urban centers. Even if the population of all the catchment areas of the capital city Addis Ababa, the only large city in

Ethiopia, is only 2.4% of the whole population, food access inequality was still high. On the other hand, small cities and its catchment areas, which cover 67% of the population and 74% in terms of area, experience the least amount of food access inequality. This large area share of communities in small cities and its catchments was the main driver of the low overall Gini index of Ethiopia compared to the low area share of communities in intermediate and large cities and their respective catchments where food access inequality was relatively high.

5.3. Limitations

While this study was successful in quantifying food access considering both physical and economic access, there is a third aspect of food access which was not considered due to the limited data referring to it: social access, which deals with cultural and consumer preference. For example, if a community's culture prefers maize but the markets in the area only sell wheat products, people can physically and economically access food but not the food within their cultural or personal preferences. With this, the FAI may characterize a community as one that has physical and economic access to food but does not identify whether the food is culturally or personally preferable for that community. Nevertheless, inclusion of social access to food is difficult to incorporate yet necessary to have an all-encompassing characterization of food access.

The construction of the FAI was dependent on the initial selection of variables. And even if the selection of the initial variables from the survey to be included in the PCA had its justifications, it was nevertheless based on personal perspective on deciding which variables were relevant. And because the index is highly dependent on the variables being put in, if a different set of variables had been selected, the constructed FAI could have been different. This is echoed by Hofstede (2001) who emphasized the importance of a priori theory in the selection of the variables to be used in a factor analysis like PCA. For example, adding more variables related to asset ownership could possibly have resulted in more variables with positive coefficients and the maximum food access index would have been higher. Although the LSMS dataset itself had a vast amount of variables, there were more variables relating to socio-economic status, specifically asset ownership, than physical access which limits the choices in terms of variables relating to different themes. More intricate selection methods could have been in place instead of a variable selection based on personal perspective on what variables to include. Moreover, further inspection of each variable to be included, specifically the remotely sensed variables, could have also been done. The insights gained from the PCA are given with caveats considering the high dependency on the selection of variables which should have been less subjective to prevent bias.

The assessment of the final GAM was done using a training-validation split of 70:30 to prevent overfitting. This is a common cross-validation approach to assess the performance of a model but there are also other approaches like k-fold cross validation as done by similar studies using similar methodological frameworks (Chi et al., 2022; Jean et al., 2016; Yeh et al., 2020). K-fold cross validation could be a better choice as samples are used in both training and validating at different folds but would also require more computing power and time. Nevertheless, the comparable performance of our final GAM to the existing models for wealth prediction in the literature could be attributed to the difference in input data. In the other studies, for example, satellite imagery, data that usually contains certain amounts of noise, was used as an input but not in our study. Furthermore, the spatial coverage of this study was only for one country while the existing studies mentioned conducted multi-country analyses.

It should also be noted that although a quantification of the number of people experiencing a certain level of food access was done, generalizations on individuals cannot be made using the FAI as majority of the data was collected at household and community level. Therefore, the generalizations made from our results are only applicable at this specific aggregated level (Schwartz, 1994). The same was emphasized by Hofstede (2001) that the manner of data collection affects the generalizability of findings from such data.

Another limitation of this study is that although the FAI is predicted at a high spatial detail for a country-level analysis, it is quite coarse for intra-urban analysis. The FAI map is only able to describe the spatial variations at larger spatial coverages, like from urban to peri urban to rural areas. As dynamics within an urban area are very complex, the FAI can only provide a general characterization of these areas and does not capture the nuances of food access within urban areas. For example, intra-urban inequalities in economic access are ever present especially in larger cities (Fotso, 2006; Haddad and Nedović-Budić, 2006) but is not clearly highlighted based on the FAI map and food access inequality results. As mentioned, this study provides community level insights and not household or individual level insights which are needed for intra-urban analysis.

5.4. Implications

Having a robust and dynamic model that quantifies food access also opened opportunities for scenario development and planning. Because the final variables used in predicting food access were easily interpretable, a scenario development application[4] was developed using the R package "shiny" v.1.6.0 (Chang et al., 2021) that allowed users to test the impact of changes in each variable to the food access index and food access inequality (see Figure 25). Additionally, providing extreme but still realistic inputs, in the form of shocks like increased cereal prices or hampered physical access due to conflict or war, also tested the robustness of the model, albeit only qualitatively. For example, the changes in food access levels, population affected, and food access inequality brought about by a hypothetical 40% increase in cereal prices can be immediately seen using the web application as shown in Figure 25. Also, the impacts of conflict towards travel time to markets, and consequently food access, can also be simulated using the application. This is especially important because Ethiopia suffered devastating impacts from the recent Tigray war where food security, specifically food access, was one of the major aspects affected (de Waal, 2021; Sheepshanks, 2022). The Integrated Food Security Phase Classification's (IPC) acute food insecurity analysis for May and September 2021 (IPC, 2021) provided information on how many people were in crisis, emergency, and catastrophe and showed where these people were located at a regional level. This scenario development application could be used to rapidly assess and even simulate future impacts of such shocks to food access at a higher spatial detail to pinpoint solutions not only at regional level but at high spatial precision. On a broader perspective, the ongoing war in Ukraine has recently made the United Nations issue a stark alert of

[4] Available at bit.ly/FoodAccessScenarios

a possible record wave of global hunger caused by price shocks and even identified countries in Sub-Saharan Africa as drastically susceptible (United Nations, 2022).

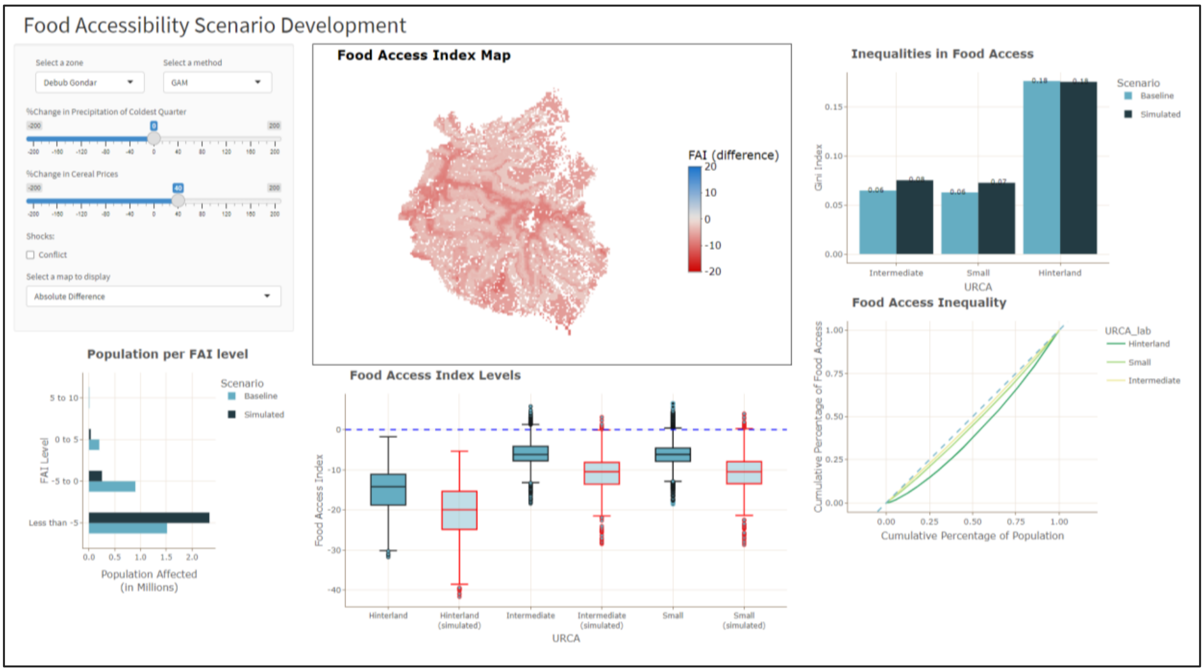


Figure 25. Screenshot of the food access scenario development application showing the impact of a ~40% increase in cereal prices towards food access.

Food access is an under-researched pillar of food security. This new FAI could be used to help inform one of the possible underlying causes of food insecurity in Ethiopia because not only does it provide the status of food access but also decomposes food access, to a degree, into physical and economic access. Additionally, it does so at a very high spatial detail of 1 km resolution, even more detail than the recently published relative wealth index by Chi et al. (2022) that provides information on asset-based wealth at 2.4 km resolution. The results from this study serves as a proof of concept that quantifying food access across space and time at high spatial detail for an entire country is possible.

5.5. Recommendations

To further ensure reliability and usability of the results, improve the quantification of food access, and gain more insights from this approach, this study has the following recommendations. First is to conduct validation of the FAI results on the ground or at least with experts in the field. Although the survey data collected acts as on the ground data, these were collected for a slightly different purpose. Another recommendation to test the applicability of the FAI is to compare with other existing indices like the RWI, as mapped by Chi et al. (2022), or other datasets like census data from local government agencies to see whether the insights gained from the FAI are in agreement.

This study implemented the conventional GAM as defined by Wood (2011) but several improvements to the conventional GAM have already been developed like boosted GAMs (Maloney et al., 2012) which have increased the predictive accuracy and have made the interpretability of the relationship between predictor and response even clearer. Apart from this, the possibility of implementing ensemble learning with a combination of different types of models can be done. Ensemble learning has produced more than satisfactory results in other applications when used in the correct context (Huang et al., 2020; Kiangala and Wang, 2021). Testing relatively more complex models like an ensemble model may improve predictive performance but may also add complications in terms of interpretability depending on the base learners used.

Lastly, since this study already provided a proof of concept for the quantification of food access at high spatial detail for an entire country, it is recommended to test this approach for multiple countries to gain global insights into food access. With this, inter-country variations in food access can be investigated apart from the within-country quantifications of food access. Also, the insights in terms of the significant factors of food access can also be tested. Food systems are very diverse, even within a country, so this diversity in food systems only increases across countries and across continents yet it is also possible to find similar food systems in countries from different continents. Capturing this variation of food access across the globe only helps in understanding our food systems, at least in terms of what drives food access.

6. CONCLUSION

This research primarily aimed to quantify food access at high spatial detail across a large spatial and temporal coverage, considering both physical and economic access, in Ethiopia through a smart integration of HH survey data and geospatial data. It also aimed to quantify inequalities in food access throughout urban-rural catchment areas. To accomplish this, three objectives were set: to develop an index from HH surveys to quantify food access, predict this index across the entire Ethiopia and across months in the year using spatiotemporal geospatial data, and finally quantify the inequalities from the predictions of the index.

This study successfully constructed the food access index from LSMS survey data. And in the process of index construction, we further discovered that while both economic and physical access to food contributed to overall food access, economic access had a stronger influence. The food access index was also effectively predicted for all populated areas in Ethiopia using a GAM trained only on four geospatial predictors. The four predictors represented physical access (travel time to markets and road proximity), economic access (road proximity as an urban-rural proxy and cereal prices as food affordability), and climate (precipitation of coldest quarter). Food access was found to be both low and temporally volatile across the year in peri-urban areas where a large proportion of the population resides. Also, temporal patterns in food access in peri-urban areas largely followed the crop calendar making these areas extremely vulnerable to shocks like drought and its cascading effects like price increases. Lastly, inequality in food access was quantified for all of Ethiopia as well as across urban-rural catchment areas using the Gini index. Food access inequality was high in peri-urban areas of large and intermediate cities but was low in small cities and its catchment areas. And because of the large share of small cities and its catchment areas compared to other urban areas, also led to the low overall food access inequality in Ethiopia.

Overall, this research served as a proof-of-concept that quantifying food access at such spatial and temporal detail and coverage is possible with existing datasets and methods. This opens up opportunities for research in the under-studied food security pillar of food access. Furthermore, the scenario development application developed in this study also provides opportunities for simulations of food access in light of possible shocks. This is timely with the ongoing war in Ukraine which brings the risks of price spikes which could lead to global hunger.

LIST OF REFERENCES

Abay, K., Hirvonen, K., 2017. Does Market Access Mitigate the Impact of Seasonality on Child Growth? Panel Data Evidence from Northern Ethiopia. J. Dev. Stud. 53, 1414–1429. https://doi.org/10.1080/00220388.2016.1251586

Abu Hatab, A., Cavinato, M.E.R., Lindemer, A., Lagerkvist, C.J., 2019. Urban sprawl, food security and agricultural systems in developing countries: A systematic review of the literature. Cities 94, 129–142. https://doi.org/10.1016/J.CITIES.2019.06.001

Archer, E., 2021. rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics. https://cran.r-project.org/package=rfPermute

Ball, K., Timperio, A., Crawford, D., 2009. Neighbourhood socioeconomic inequalities in food access and affordability. Health Place 15, 578–585. https://doi.org/10.1016/J.HEALTHPLACE.2008.09.010

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115. https://doi.org/10.1016/J.INFFUS.2019.12.012

Barrett, C.B., 2010. Measuring Food Insecurity. Science (80-.). 327, 825–828. https://doi.org/10.1126/science.1182768

Battersby, S.E., Finn, M.P., Usery, E.L., Yamamoto, K.H., 2014. Implications of web mercator and its use in online mapping. Cartographica 49, 85–101. https://doi.org/10.3138/carto.49.2.2313

Bell, W., Lividini, K., Masters, W.A., 2021. Global dietary convergence from 1970 to 2010 altered inequality in agriculture, nutrition and health. Nat. Food 2, 156–165. https://doi.org/10.1038/s43016-021-00241-9

Bono, F., Finn, J.C., 2016. Food Diaries to Measure Food Access: A Case Study from Rural Cuba. Prof. Geogr. 69, 59–69. https://doi.org/10.1080/00330124.2016.1157499

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Breyer, B., Voss-Andreae, A., 2013. Food mirages: Geographic and economic barriers to healthful food access in Portland, Oregon. Health Place 24, 131–139. https://doi.org/10.1016/J.HEALTHPLACE.2013.07.008

Brown, M.E., 2016. Remote sensing technology and land use analysis in food security assessment. J. Land Use Sci. 11, 623–641. https://doi.org/10.1080/1747423X.2016.1195455

Burgert, C.R., Colston, J., Roy, T., Zachary, B., 2013. Geographic Displacement Procedure and Georeferenced Data Release Policy for the Demographic and Health Surveys. DHS Spatial Analysis Reports No. 7. https://dhsprogram.com/pubs/pdf/SAR7/SAR7.pdf

Cafiero, C., Nord, M., Viviani, S., Del Grossi, M.E., Ballard, T.J., Kepple, A., Miller, M., Nwosu, C., 2016. Methods for Estimating Comparable Prevalence Rates of Food Insecurity Experienced by Adults Throughout the World. http://www.fao.org/3/a-i4830e.pdf

Campbell, C.C., 1991. Food Insecurity: A Nutritional Outcome or a Predictor Variable? J. Nutr. 121, 408–415. https://doi.org/10.1093/jn/121.3.408

Cattaneo, A., Adukia, A., Brown, D.L., Christiaensen, L., Evans, D.K., Haakenstad, A., McMenomy, T., Partridge, M., Vaz, S., Weiss, D.J., 2022. Economic and social development along the urban–rural continuum: New opportunities to inform policy. World Dev. 157, 105941. https://doi.org/10.1016/J.WORLDDEV.2022.105941

Cattaneo, A., Adukia, A., Brown, D.L., Christiaensen, L., Evans, D.K., Haakenstad, A., McMenomy, T., Patridge, M., Vaz, S., Weiss, D.J., 2021a. Economic and Social Development along the Urban-Rural Continuum, Policy Research Working Paper (No. 9756), World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/36199

Cattaneo, A., Nelson, A., McMenomy, T., 2021b. Global mapping of urban–rural catchment areas reveals unequal access to services. Proc. Natl. Acad. Sci. 118. https://doi.org/10.1073/PNAS.2011990118

Cedrez, C.B., Chamberlin, J., Guo, Z., Hijmans, R.J., 2020a. Spatial variation in fertilizer prices in Sub-Saharan Africa. PLoS One 15, e0227764. https://doi.org/10.1371/JOURNAL.PONE.0227764

Cedrez, C.B., Chamberlin, J., Hijmans, R.J., 2020b. Seasonal, annual, and spatial variation in cereal prices in Sub-Saharan Africa. Glob. Food Sec. 26. https://doi.org/10.1016/J.GFS.2020.100438

Central Statistics Agency of Ethiopia, 2020. Ethiopia Socioeconomic Survey (ESS4) 2018-2019. Public Use Dataset. Ref: ETH_2018_ESS_v01. [Dataset]. URL https://microdata.worldbank.org/index.php/catalog/3823 (accessed 5.18.21).

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2021. shiny: Web Application Framework for R. https://cran.r-project.org/package=shiny
Chen, X., 2017. Take the edge off: A hybrid geographic food access measure. Appl. Geogr. 87, 149–159. https://doi.org/10.1016/J.APGEOG.2017.07.013
Chi, G., Fang, H., Chatterjee, S., Blumenstock, J.E., 2022. Microestimates of wealth for all low- and middle-income countries. Proc. Natl. Acad. Sci. U. S. A. 119. https://doi.org/10.1073/PNAS.2113658119
Cowell, F., 2009. Measuring Inequality, Third ed., LSE Perspectives in Economic Analysis. Oxford Univeristy Press, New York. https://doi.org/10.1093/acprof:osobl/9780199594030.001.0001
D'Odorico, P., Carr, J.A., Davis, K.F., Dell'Angelo, J., Seekell, D.A., 2019. Food Inequality, Injustice, and Rights. Bioscience 69, 180–190. https://doi.org/10.1093/BIOSCI/BIZ002
de Waal, A., 2021. Viewpoint: Why Ethiopia's Tigray region is starving, but no famine declared [WWW Document]. BBC News. URL https://www.bbc.com/news/world-africa-58921744 (accessed 4.29.22).
Deller, S., Canto, A., Brown, L., 2015. Rural poverty, health and food access. Reg. Sci. Policy Pract. 7, 61–74. https://doi.org/10.1111/RSP3.12056
Díaz-Uriarte, R., Alvarez de Andrés, S., 2006. Gene selection and classification of microarray data using random forest. BMC Bioinformatics 7. https://doi.org/10.1186/1471-2105-7-3
Downs, S.M., Fox, E.L., 2021. Uneven decline in food system inequality. Nat. Food 2, 141–142. https://doi.org/10.1038/s43016-021-00247-3
Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Stat. Sci. 11, 89–121. https://doi.org/10.1214/SS/1038425655
Elbers, C., Fujii, T., Lanjouw, P., Özler, B., Yin, W., 2007. Poverty alleviation through geographic targeting: How much does disaggregation help? J. Dev. Econ. 83, 198–213. https://doi.org/10.1016/j.jdeveco.2006.02.001
Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., E. Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography (Cop.). 29, 129–151. https://doi.org/10.1111/J.2006.0906-7590.04596.X
Ezeh, A., Kissling, F., Singer, P., 2020. Why sub-Saharan Africa might exceed its projected population size by 2100. Lancet 396, 1131–1133. https://doi.org/10.1016/S0140-6736(20)31522-1
Fanzo, J., Haddad, L., Schneider, K.R., Béné, C., Covic, N.M., Guarin, A., Herforth, A.W., Herrero, M., Sumaila, U.R., Aburto, N.J., Amuyunzu-Nyamongo, M., Barquera, S., Battersby, J., Beal, T., Bizzotto Molina, P., Brusset, E., Cafiero, C., Campeau, C., Caron, P., Cattaneo, A., Conforti, P., Davis, C., DeClerck, F.A.J., Elouafi, I., Fabi, C., Gephart, J.A., Golden, C.D., Hendriks, S.L., Huang, J., Laar, A., Lal, R., Lidder, P., Loken, B., Marshall, Q., Masuda, Y.J., McLaren, R., Neufeld, L.M., Nordhagen, S., Remans, R., Resnick, D., Silverberg, M., Torero Cullen, M., Tubiello, F.N., Vivero-Pol, J.-L., Wei, S., Rosero Moncayo, J., 2021. Viewpoint: Rigorous monitoring is necessary to guide food system transformation in the countdown to the 2030 global goals. Food Policy 104, 102163. https://doi.org/10.1016/J.FOODPOL.2021.102163
FAO, 2021a. FAOSTAT Monthly Food CPI and General CPI Database [Dataset]. https://www.fao.org/faostat/en/#data/CP (accessed 5.4.22).
FAO, 2021b. GIEWS Country Brief: Ethiopia [WWW Document]. URL https://www.fao.org/giews/countrybrief/country/ETH/pdf/ETH.pdf (accessed 6.8.22)
FAO, 1996. Rome Declaration on World Food Security and World Food Summit Plan of Action, in: World Food Summit. Rome, Italy. http://www.fao.org/docrep/003//w3613e/w3613e00.htm
FAO, IFAD, UNICEF, WFP, WHO, 2021. The State of Food Security and Nutrition in the World 2021, Transforming food systems for food security, improved nutrition and affordable healthy diets for all. Rome, FAO. https://doi.org/10.4060/cb5409en
FAO, IFAD, UNICEF, WFP, WHO, 2018. The State of Food Security and Nutrition in the World 2018, Building climate resilience for food security and nutrition. Rome, FAO. https://www.wfp.org/publications/2018-state-food-security-and-nutrition-world-sofi-report
FAO, IIASA, 2021. Global Agro Ecological Zones version 4 (GAEZ v4) [Dataset]. URL

http://www.fao.org/gaez/ (accessed 2.21.22).
Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15, 3133–3181. https://doi.org/10.5555/2627435.2697065
Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315. https://doi.org/10.1002/JOC.5086
Filmer, D., Pritchett, L., 2001. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. Demography 38, 115–132. https://doi.org/10.1353/DEM.2001.0003
Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., Zaks, D.P.M., 2011. Solutions for a cultivated planet. Nature 478, 337–342. https://doi.org/10.1038/nature10452
Fotso, J.-C., 2007. Urban-rural differentials in child malnutrition: trends and socioeconomic correlates in sub-Saharan Africa. Health Place 13, 205–223. https://doi.org/10.1016/J.HEALTHPLACE.2006.01.004
Fotso, J.C., 2006. Child health inequities in developing countries: Differences across urban and rural areas. Int. J. Equity Health 5. https://doi.org/10.1186/1475-9276-5-9
Frelat, R., Lopez-Ridaura, S., Giller, K.E., Herrero, M., Douxchamps, S., Djurfeldt, A.A., Erenstein, O., Henderson, B., Kassie, M., Paul, B.K., Rigolot, C., Ritzema, R.S., Rodriguez, D., Van Asten, P.J.A., Van Wijk, M.T., 2016. Drivers of household food availability in sub-Saharan Africa based on big data from small farms. Proc. Natl. Acad. Sci. U. S. A. 113, 458–463. https://doi.org/10.1073/pnas.1518384112
Garcia, X., Garcia-Sierra, M., Domene, E., 2020. Spatial inequality and its relationship with local food environments: The case of Barcelona. Appl. Geogr. 115, 102140. https://doi.org/10.1016/J.APGEOG.2019.102140
Genuer, R., Poggi, J.M., Tuleau-Malot, C., 2010. Variable selection using random forests. Pattern Recognit. Lett. 31, 2225–2236. https://doi.org/10.1016/J.PATREC.2010.03.014
Gini, C., 1936. On the Measure of Concentration with Special Reference to Income and Statistics, in: Colorado College Publication, General Series No. 208. pp. 73–79.
Graham, E., Miller, M., Viviani, S., 2014. Piloting the Food Insecurity Experience Scale (FIES) in Africa, south of the Sahara: Assessing food insecurity severity to support policy solutions. Nat. Faune 28, 49–53. https://www.fao.org/publications/card/en/c/8a2a4982-9578-4cf9-a43a-5f6d334e20e0/
Grosh, M.E., Glewwe, P., Duncan, G.J., 1998. Data Watch: The World Bank's Living Standards Measurement Study Household Surveys. J. Econ. Perspect. 12, 187–196. https://doi.org/10.1257/JEP.12.1.187
Haddad, M.A., Nedović-Budić, Z., 2006. Using Spatial Statistics to Analyze Intra-urban Inequalities and Public Intervention in São Paulo, Brazil. J. Hum. Dev. 7, 85–109. https://doi.org/10.1080/14649880500502102
Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate Data Analysis, 5th ed. Prentice Hall.
Hastie, T., 2020. gam: Generalized Additive Models. https://cran.r-project.org/package=gam
Hastie, T., Tibshirani, R., 1986. Generalized additive models. Stat. Sci. 1, 297–310. https://doi.org/10.1214/SS/1177013604
Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., Neal, J., 2022. A 30 m global map of elevation with forests and buildings removed. Environ. Res. Lett. 17, 024016. https://doi.org/10.1088/1748-9326/AC4D4F
Headey, D., Alderman, H., 2019. The Relative Caloric Prices of Healthy and Unhealthy Foods Differ Systematically across Income Levels and Continents. J. Nutr. 149, 2020–2033. https://doi.org/10.1093/JN/NXZ158
Headey, D., Hirvonen, K., Hoddinott, J., Stifel, D., 2019. Rural Food Markets and Child Nutrition. Am. J. Agric. Econ. 101, 1311–1327. https://doi.org/10.1093/AJAE/AAZ032
Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/PEERJ.5518
Hill, R., Fuje, H., 2020. What is the Impact of Weather Shocks on Prices? Evidence from Ethiopia, Policy Research Working Paper (No. 9389), World Bank, Washington, DC.

https://openknowledge.worldbank.org/handle/10986/34475
Hirvonen, K., Taffesse, A.S., Worku Hassen, I., 2016. Seasonality and household diets in Ethiopia. Public Health Nutr. 19, 1723–1730. https://doi.org/10.1017/S1368980015003237
Hofstede, G., 2001. Culture's Consequences : Comparing Values, Behaviors, Institutions, and Organizations across Nations, First ed., SAGE Publications, Thousand Oaks, California.
Hotelling, H., 1933. Analysis of a complex of statistical variables into Principal Components. J. Educ. Psychol. 24, 417–441. https://doi.org/10.1037/h0071325
Huang, J.C., Tsai, Y.C., Wu, P.Y., Lien, Y.H., Chien, C.Y., Kuo, C.F., Hung, J.F., Chen, S.C., Kuo, C.H., 2020. Predictive modeling of blood pressure during hemodialysis: a comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. Comput. Methods Programs Biomed. 195, 105536. https://doi.org/10.1016/J.CMPB.2020.105536
Husak, G.J., Marshall, M.T., Michaelsen, J., Pedreros, D., Funk, C., Galu, G., 2008. Crop area estimation using high and medium resolution satellite imagery in areas with complex topography. J. Geophys. Res. Atmos. 113, 14112. https://doi.org/10.1029/2007JD009175
ILO, IMF, OECD, UNECE, Eurostat, World Bank, 2004. Consumer price index manual: Theory and practice, International Labour Office, Geneva. https://www.ilo.org/public/english/bureau/stat/download/cpi/cpi_manual_en.pdf
International Food Policy Research Institute, 2020. Spatially-Disaggregated Crop Production Statistics Data in Africa South of the Sahara for 2017 [Dataset]. https://doi.org/10.7910/DVN/FSSKBW
IPC, 2021. IPC Acute Food Insecurity Analysis (May - September 2021) [WWW Document]. URL https://www.ipcinfo.org/fileadmin/user_upload/ipcinfo/docs/IPC_Ethiopia_Acute_Food_Insecurity_2021MaySept_national.pdf (accessed 5.25.22).
Janda, K.M., Salvo Dominguez, D., Ranjit, N., Hoelscher, D.M., Price, A., van den Berg, A., 2021. Mapping Food Insecurity-Related 2-1-1 Calls in a 10-County Area of Central Texas by Zip Code: Exploring the Role of Geographic Food Access, Urbanicity and Demographic Indicators. J. Community Health 46, 86–97. https://doi.org/10.1007/S10900-020-00847-3/TABLES/3
Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science (80-.). 353, 790–794. https://doi.org/10.1126/SCIENCE.AAF7894
Jeong, J., Liu, C.Y., 2020. Neighborhood Diversity and Food Access in a Changing Urban Spatial Structure. City Community 19, 963–986. https://doi.org/10.1111/cico.12426
Jiao, J., Moudon, A. V., Ulmer, J., Hurvitz, P.M., Drewnowski, A., 2012. How to identify food deserts: Measuring physical and economic access to supermarkets in King County, Washington. Am. J. Public Health 102, 32–39. https://doi.org/10.2105/AJPH.2012.300675
Jin, H., Lu, Y., 2021. SAR-Gi*: Taking a spatial approach to understand food deserts and food swamps. Appl. Geogr. 134, 102529. https://doi.org/10.1016/J.APGEOG.2021.102529
Karthikeyan, L., Chawla, I., Mishra, A.K., 2020. A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. J. Hydrol. 586, 1–22. https://doi.org/10.1016/J.JHYDROL.2020.124905
Kiangala, S.K., Wang, Z., 2021. An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment. Mach. Learn. with Appl. 4, 100024. https://doi.org/10.1016/J.MLWA.2021.100024
Kolak, M., Bradley, M., Block, D.R., Pool, L., Garg, G., Toman, C.K., Boatright, K., Lipiszko, D., Koschinsky, J., Kershaw, K., Carnethon, M., Isakova, T., Wolf, M., 2018. Urban foodscape trends: Disparities in healthy food access in Chicago, 2007–2014. Health Place 52, 231–239. https://doi.org/10.1016/J.HEALTHPLACE.2018.06.003
Lawrence, S., Liu, Q., Yakovenko, V.M., 2013. Global Inequality in Energy Consumption from 1980 to 2010. Entropy 15, 5565–5579. https://doi.org/10.3390/E15125565
Lê, Q., Auckland, S., Nguyen, H.B., Murray, S., Long, G., Terry, D.R., 2015. The socio-economic and physical contributors to food insecurity in a rural community. SAGE Open 5. https://doi.org/10.1177/2158244014567401
Lê, S., Josse, J., Husson, F., 2008. FactoMineR: An R Package for Multivariate Analysis. J. Stat. Softw. 25, 1–18. https://doi.org/10.18637/JSS.V025.I01
Lele, U., Masters, W.A., Kinabo, J., Meenakshi, J. V., Ramaswami, B., Tagwireyi, J., Bell, W.F.L., Goswami, S., 2016. Measuring Food and Nutrition Security: An Independent Technical Assessment and User's Guide for Existing Indicators, Measuring Food and Nutrition Security Technical

Working Group. Food Security Information Network, Rome, Italy. https://sites.tufts.edu/willmasters/files/2016/06/FSIN-TWG_UsersGuide_12June2016.pdf
Lentz, E.C., Michelson, H., Baylis, K., Zhou, Y., 2019. A data-driven approach improves food insecurity crisis prediction. World Dev. 122, 399–409. https://doi.org/10.1016/j.worlddev.2019.06.008
Li, J., Song, G., Semakula, H.M., Dou, Y., Zhang, S., 2019. Food access inequalities in Chinese urban neighborhoods: a case study of the Dalian development zone. Food Secur. 11, 1087–1099. https://doi.org/10.1007/S12571-019-00963-6/FIGURES/6
Liu, J., Shively, G.E., Binkley, J.K., 2014. Access to variety contributes to dietary diversity in China. Food Policy 49, 323–331. https://doi.org/10.1016/J.FOODPOL.2014.09.007
Liu, Y., Ke, X., Wu, W., Zhang, M., Fu, X., Li, J., Jiang, J., He, Y., Zhou, C., Li, W., Li, Y., Song, Y., Zhou, X., 2022. Geospatial characterization of rural settlements and potential targets for revitalization by geoinformation technology. Sci. Rep. 12, 1–16. https://doi.org/10.1038/s41598-022-12294-2
Lohani, S., Singh, R.B., Lohani, J., 2010. Universal primary education in Nepal: Fulfilling the right to education. Prospects 40, 355–374. https://doi.org/10.1007/S11125-010-9162-6
Lorenz, M.O., 1905. Methods of Measuring the Concentration of Wealth. Publ. Am. Stat. Assoc. 9, 209. https://doi.org/10.2307/2276207
Macintyre, S., 2007. Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? Int. J. Behav. Nutr. Phys. Act. 4, 1–7. https://doi.org/10.1186/1479-5868-4-32
Maloney, K.O., Schmid, M., Weller, D.E., 2012. Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. Methods Ecol. Evol. 3, 116–128. https://doi.org/10.1111/j.2041-210X.2011.00124.x
Marshall, M., Norton-Griffiths, M., Herr, H., Lamprey, R., Sheffield, J., Vagen, T., Okotto-Okotto, J., 2017. Continuous and consistent land use/cover change estimates using socio-ecological data. Earth Syst. Dyn. 8, 55–73. https://doi.org/10.5194/ESD-8-55-2017
Menchini, L., Redmond, G., 2009. Poverty and deprivation among children in Eastern Europe and Central Asia. Int. J. Soc. Welf. 18, 225–236. https://doi.org/10.1111/j.1468-2397.2008.00620.x
Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. Ecol. Modell. 411, 108815. https://doi.org/10.1016/J.ECOLMODEL.2019.108815
Mohammed, I., Marshall, M., de Bie, K., Estes, L., Nelson, A., 2020. A blended census and multiscale remote sensing approach to probabilistic cropland mapping in complex landscapes. ISPRS J. Photogramm. Remote Sens. 161, 233–245. https://doi.org/10.1016/J.ISPRSJPRS.2020.01.024
Molnar, C., 2019. Interpretable machine learning. A Guide for Marking Black Box Models Explainable. https://christophm.github.io/interpretable-ml-book/.
Mutanga, O., Dube, T., Galal, O., 2017. Remote sensing of crop health for food security in Africa: Potentials and constraints. Remote Sens. Appl. Soc. Environ. 8, 231–239. https://doi.org/10.1016/J.RSASE.2017.10.004
Nelson, A., Weiss, D.J., van Etten, J., Cattaneo, A., McMenomy, T.S., Koo, J., 2019. A suite of global accessibility indicators. Sci. Data 6, 1–9. https://doi.org/10.1038/s41597-019-0265-5
Nyangasa, M.A., Buck, C., Kelm, S., Sheikh, M., Hebestreit, A., 2019. Exploring Food Access and Sociodemographic Correlates of Food Consumption and Food Insecurity in Zanzibari Households. Int. J. Environ. Res. Public Health 16, 1557. https://doi.org/10.3390/IJERPH16091557
Olana, G., Kedir, N., Abagodu, R., Dasgupta, B., Ambelu, W., Okello, F., Magut, M., 2018. Crop Availability and Market Analysis in Ethiopia. https://www.usaid.gov/sites/default/files/documents/1866/Crop_Availability_and_Market_Analysis_in_Ethiopia_05.09.18_508_Compliant_1.pdf
Park, N.W., Chi, K.H., 2008. Quantitative assessment of landslide susceptibility using high-resolution remote sensing data and a generalized additive model. Int. J. Remote Sens. 29, 247–264. https://doi.org/10.1080/01431160701227661
Plaut, M., 2008. Ethiopian millions "risk hunger" [WWW Document]. BBC News. URL http://news.bbc.co.uk/2/hi/africa/7409623.stm (accessed 5.20.22).
Pozzi, F., Robinson, T., 2008. Accessibility Mapping in the Horn of Africa: Applications for Livestock Policy, IGAD LPI Working Paper 11-08. https://cgspace.cgiar.org/handle/10568/24984
Radke, J., Mu, L., 2009. Spatial Decompositions, Modeling and Mapping Service Regions to Predict Access to Social Programs. Geogr. Inf. Sci. 6, 105–112. https://doi.org/10.1080/10824000009480538

Rein, M., Miller, S.M., 1974. Standards of Income Redistribution. Challenge 17, 20–26. https://doi.org/10.1080/05775132.1974.11470052
Rutstein, S.O., Johnson, K., 2004. The DHS Wealth Index, DHS Comparative Reports No. 6. Calverton, Maryland: ORC Macro. https://dhsprogram.com/publications/publication-cr6-comparative-reports.cfm
Schwartz, S., 1994. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. Am. J. Public Health 84, 819–824. https://doi.org/10.2105/AJPH.84.5.819
Seekell, D.A., D'Odorico, P., Pace, M.L., 2011. Virtual water transfers unlikely to redress inequality in global water use. Environ. Res. Lett. 6. https://doi.org/10.1088/1748-9326/6/2/024017
Sen, A., 1981. Poverty and Famines: An Essay on Entitlement and Deprivation. Oxford Univeristy Press, New York. https://www.prismaweb.org/nl/wp-content/uploads/2017/06/Poverty-and-famines┃Amartya-Sen┃1981.pdf
Seyoum Taffesse, A., Dorosh, P., Gemessa, S.A., 2011. Crop production in Ethiopia: Regional patterns and trends (No. 0016), Ethiopia Strategy Support Program II. https://doi.org/10.9783/9780812208610.53
Sheepshanks, O., 2022. Man-made Famine in Tigray – The World's Worst Hunger Crisis in a Decade [WWW Document]. Omna Tigray. URL https://omnatigray.org/man-made-famine-in-tigray-the-worlds-worst-hunger-crisis-in-a-decade/ (accessed 4.29.22).
Sishodia, R.P., Ray, R.L., Singh, S.K., 2020. Applications of Remote Sensing in Precision Agriculture: A Review. Remote Sens. 12, 3136. https://doi.org/10.3390/RS12193136
UN General Assembly, 2017. Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development, A/RES/71/3. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N17/207/63/PDF/N1720763.pdf
UN General Assembly, 2015. Transforming our world : the 2030 Agenda for Sustainable Development, A/RES/70/1, 1–35. https://www.refworld.org/docid/57b6e3e44.html
UNDP, 2020. Technical Notes, Human Development Reports. http://hdr.undp.org/sites/default/files/hdr2020_technical_notes.pdf
United Nations, 2022. War in Ukraine threatens to unleash "unprecedented wave" of global hunger and destitution, warns UN Chief [WWW Document]. Sustain. Dev. Goals Blog. URL https://www.un.org/sustainabledevelopment/blog/2022/06/war-in-ukraine-threatens-to-unleash-unprecedented-wave-of-global-hunger-and-destitution-warns-un-chief/ (accessed 6.10.22).
Usman, M.A., Callo-Concha, D., 2021. Does market access improve dietary diversity and food security? Evidence from Southwestern Ethiopian smallholder coffee producers. Agric. Food Econ. 9. https://doi.org/10.1186/S40100-021-00190-8
Usman, M.A., Haile, M.G., 2019. Market access, food security and nutrition: evidence from selected sub-Saharan African Countries, in: 6th African Conference of Agricultural Economists. pp. 1–23. https://doi.org/10.22004/ag.econ.295778
van Etten, J., 2017. R Package gdistance : Distances and Routes on Geographical Grids. J. Stat. Softw. 76. https://doi.org/10.18637/jss.v076.i13
Velazco, J., Ballester, R., 2015. Food Access and Shocks in Rural Households: Evidence from Bangladesh and Ethiopia. Soc. Indic. Res. 129, 527–549. https://doi.org/10.1007/S11205-015-1118-Z
Wang, H., Liu, C., Fan, H., Tian, X., 2017. Rising food accessibility contributed to the increasing dietary diversity in rural and urban China. Asia Pac. J. Clin. Nutr. 26, 738–747. https://doi.org/10.6133/apjcn.052016.03
Weiss, D.J., Nelson, A., Gibson, H.S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T.C.D., Howes, R.E., Tusting, L.S., Kang, S.Y., Cameron, E., Bisanzio, D., Battle, K.E., Bhatt, S., Gething, P.W., 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature 553, 333–336. https://doi.org/10.1038/nature25181
Weiss, D.J., Nelson, A., Vargas-Ruiz, C.A., Gligorić, K., Bavadekar, S., Gabrilovich, E., Bertozzi-Villa, A., Rozier, J., Gibson, H.S., Shekel, T., Kamath, C., Lieber, A., Schulman, K., Shao, Y., Qarkaxhija, V., Nandi, A.K., Keddie, S.H., Rumisha, S., Amratia, P., Arambepola, R., Chestnutt, E.G., Millar, J.J., Symons, T.L., Cameron, E., Battle, K.E., Bhatt, S., Gething, P.W., 2020. Global maps of travel time to healthcare facilities. Nat. Med. 26, 1835–1838. https://doi.org/10.1038/s41591-020-1059-1
Westerveld, J.J.L., van den Homberg, M.J.C., Nobre, G., van den Berg, D.L.J., Teklesadik, A. D., Stuit, S. M., 2021. Forecasting transitions in the state of food security with machine learning using transferable features. Sci. Total Environ. 786, 1–15.

https://doi.org/10.1016/j.scitotenv.2021.147366
Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York. https://ggplot2.tidyverse.org
Wood, S.N., 2017. Generalized additive models: An introduction with R, Second ed., Chapman and Hall, New York. https://doi.org/10.1201/9781315370279
Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B (Statistical Methodol. 73, 3–36. https://doi.org/10.1111/J.1467-9868.2010.00749.X
Wood, S.N., 2003. Thin plate regression splines. J. R. Stat. Soc. Ser. B (Statistical Methodol. 65, 95–114. https://doi.org/10.1111/1467-9868.00374
World Food Programme, 2021a. Global Market and Economic Analysis Database [Dataset]. URL https://dataviz.vam.wfp.org/economic_explorer/price-forecasts-alerts?adm0=79 (accessed 9.14.21).
World Food Programme, 2021b. Global Food Prices Database [Dataset]. URL https://data.humdata.org/dataset/wfp-food-prices (accessed 2.2.22).
WorldPop, Center for International Earth Science Information Network, Columbia University, 2018. Ethiopia Population Density 2018 UN Adjusted [Dataset]. URL https://doi.org/10.5258/SOTON/WP00675 (accessed 1.12.22)
Wu, W., Yu, Q., You, L., Chen, K., Tang, H., Liu, J., 2018. Global cropping intensity gaps: Increasing food production without cropland expansion. Land use policy 76, 515–525. https://doi.org/10.1016/J.LANDUSEPOL.2018.02.032
Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., Burke, M., 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. Nat. Commun. 11, 1–11. https://doi.org/10.1038/s41467-020-16185-w
Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E., Ramoino, F., Arino, O., 2021. ESA WorldCover 10m 2020 v100 (Version 100) [Dataset]. https://doi.org/10.5281/zenodo.5571936 (accessed 10.20.22)
Zezza, A., Tasciotti, L., 2010. Urban agriculture, poverty, and food security: Empirical evidence from a sample of developing countries. Food Policy 35, 265–273. https://doi.org/10.1016/J.FOODPOL.2010.04.007
Zhang, Y., Xie, Y., Li, L., 2012. Crash frequency analysis of different types of urban roadway segments using generalized additive model. J. Safety Res. 43, 107–114. https://doi.org/10.1016/J.JSR.2012.01.003

7. ANNEX

7.1. Spatiotemporal food price prediction

For the spatiotemporal prediction of the FAI, spatiotemporal data for food prices also had to be prepared apart from existing spatial data. The methodology in Cedrez et al. (2020b) which looked at the spatial and temporal differences of cereal prices across sub-Saharan Africa, was adopted and slightly modified.

7.1.1. Market food prices pre-processing

To create the 1 km resolution raster for food prices, first, the global WFP VAM food price dataset was filtered to only prices in Ethiopia then the different crops were grouped into six major crop groups: cereals, pulses, oilseeds, vegetables, fruits, and tubers. Only the crop group of cereals was used in the next steps because there was not enough data across different markets and across time for the other crop groups.

For prices to be comparable across time, all prices were adjusted for inflation by dividing them by the monthly CPI. In Cedrez et al. (2020), they also adjusted for inter-country differences by dividing the prices with the purchasing power parity (PPP) but this was not necessary in this study since the study area is only a single country: Ethiopia. Furthermore, the food price data contained both retail and wholesale prices for some markets. Similarly, to have comparable price ranges within the dataset, wholesale prices were converted to retail prices by fitting linear regression models using price data from markets that had both retail and wholesale prices, as done in Cedrez et al. (2020). Food prices in markets with only retail prices were left as is. The final adjusted cereal prices were plotted across the years from 2000 to 2021 as well as across months within a year.

Variables relating to demand, market access, and environmental factors affecting crop production were used to estimate cereal price across space for all of Ethiopia across different months. We slightly modified the method by Cedrez et al. (2020a) where they used similar datasets to predict fertilizer prices. In their study, they used latitude and longitude to represent location but this was not done in this study because Meyer et al. (2019) have cautioned the usage of latitude and longitude as predictor variables as these tend to lead to overfitting and therefore unreliable predictions on unseen data. And when both latitude and longitude have the largest significance, spatial outputs tend to show banding, where predictions will be very similar along a straight line, across a certain latitude or longitude value, which is not natural in this case. To represent demand, population density data from WorldPop was used while market access was represented by travel time to the nearest town with a population between (1) 20,000 and 50,000, (2) 200,000 and 500,000, and (3) 1 million and 5 million. Monthly precipitation data from WorldClim (Fick and Hijmans, 2017) represented environmental factors.

After grouping crops by crop group, there was only sufficient price data for the cereal price group. And although pulses had data in many unique markets as well, some of these were only for early years between 2000-2005 while cereal prices were relatively consistent across time and across markets. Overall, there was food price data in 110 unique markets and Table 6 shows the breakdown of available data per crop group.

Table 6. Summary of grouped crop price data.

Crop group	Number of unique markets with data	Consistent temporal data
Cereals	105	Yes
Pulses	80	No
Oil seeds	14	Yes

Vegetables	16	Yes
Fruits	8	Yes
Tubers	17	Yes

The annual average cereal prices per year were plotted after deflating cereal prices for temporal differences using monthly CPI data. Figure 26 shows a clear spike in average cereal prices in 2008 which represent the effects of the drought that occurred in the two previous planting seasons (Plaut, 2008).

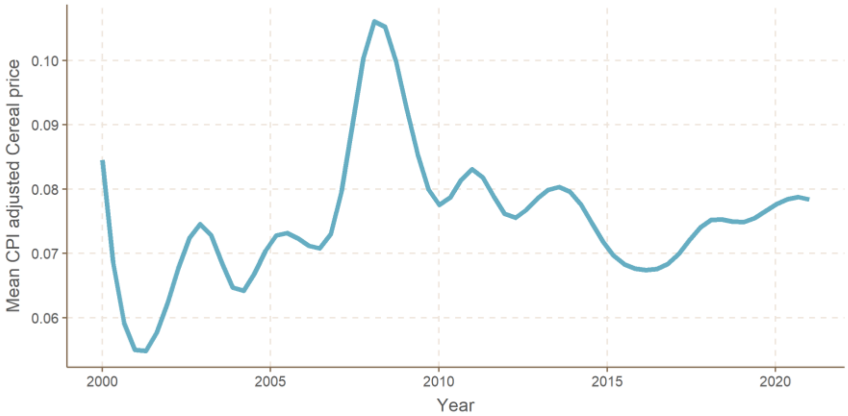


Figure 26. Mean CPI-adjusted cereal price from 2000 to 2021.

In the linear regression model for estimating the relationship between wholesale and retail prices, wholesale prices for cereals was estimated to be 0.8 times the retail prices based on the slope coefficient in the linear regression. The model had an adjusted R^2 of 0.14 and the fit of the wholesale and retail prices are seen in Figure 27. The total number of unique market-month-year price records was 7,408.

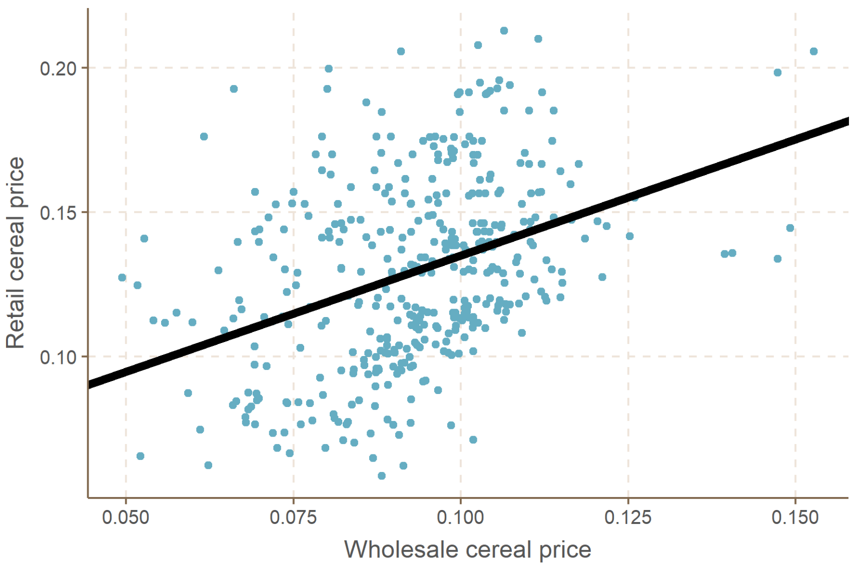


Figure 27. Linear regression fit of retail and wholesale cereal price.

7.1.2. Cereal price prediction using random forest

Using a training and validation split of 70:30, a RF model was tuned and trained using these five variables to predict monthly cereal price where only precipitation was changed for each respective month and all other variables were assumed to be unchanged. The performance of the cereal price prediction model was assessed using both the out-of-bag (OOB) samples internally determined by the RF model and the validation set. The mean OOB error and mean R^2 were calculated after 100 repetitions to deal with the random nature of the RF. Also, two metrics were calculated from the validation set: the RMSE and nRMSE. The scatterplot of predicted against observed values for the validation set along with the aforementioned metrics are seen in Figure 28. The trained RF model was then used to predict the annual average cereal price for the entire Ethiopia as well as for all 12 months.

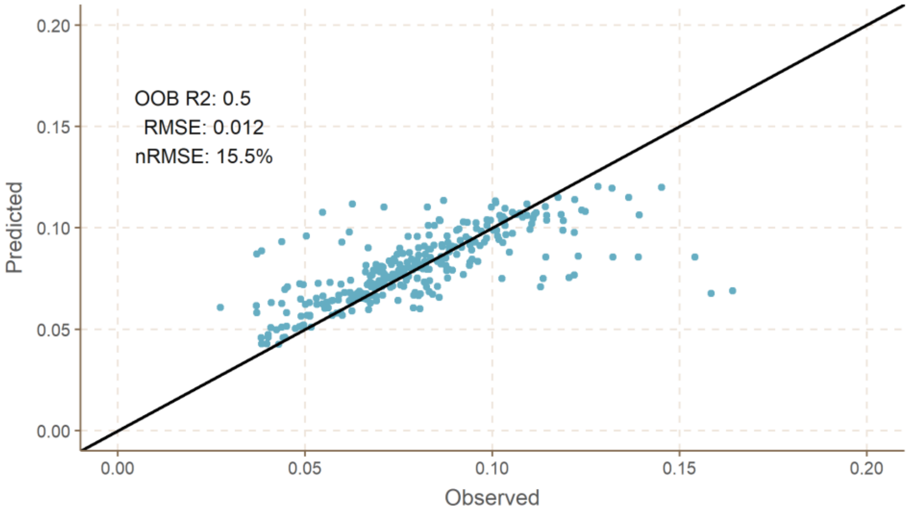


Figure 28. Scatterplot of cereal price prediction results with performance metrics.

Annual average CPI-adjusted cereal price was predicted throughout populated areas in Ethiopia using RF. The central urban area of Ethiopia experiences relatively low cereal prices compared to the outskirt rural areas which experience almost double the prices especially in the south eastern region of Somali. Figure 29 shows the map of predicted cereal price in Ethiopia.

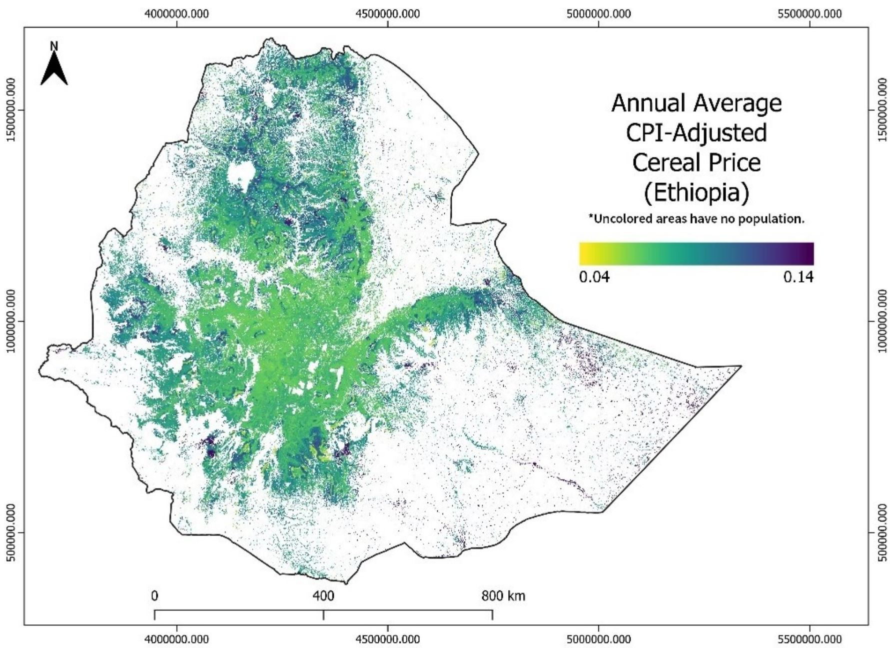


Figure 29. Map of annual average cereal price in Ethiopia.

With only precipitation changing for each month, the monthly cereal price was predicted using the trained model that predicted the annual average. The monthly coefficient of variation of cereal price from the annual average was calculated and mapped in Figure 30. Highest prices in the central urban area of Ethiopia occur during July and August while prices are generally low from November to January while for the hinterland areas, the northwest region of Ethiopia experiences high cereal prices from December to February. The difference in prices during the other months were not as high.

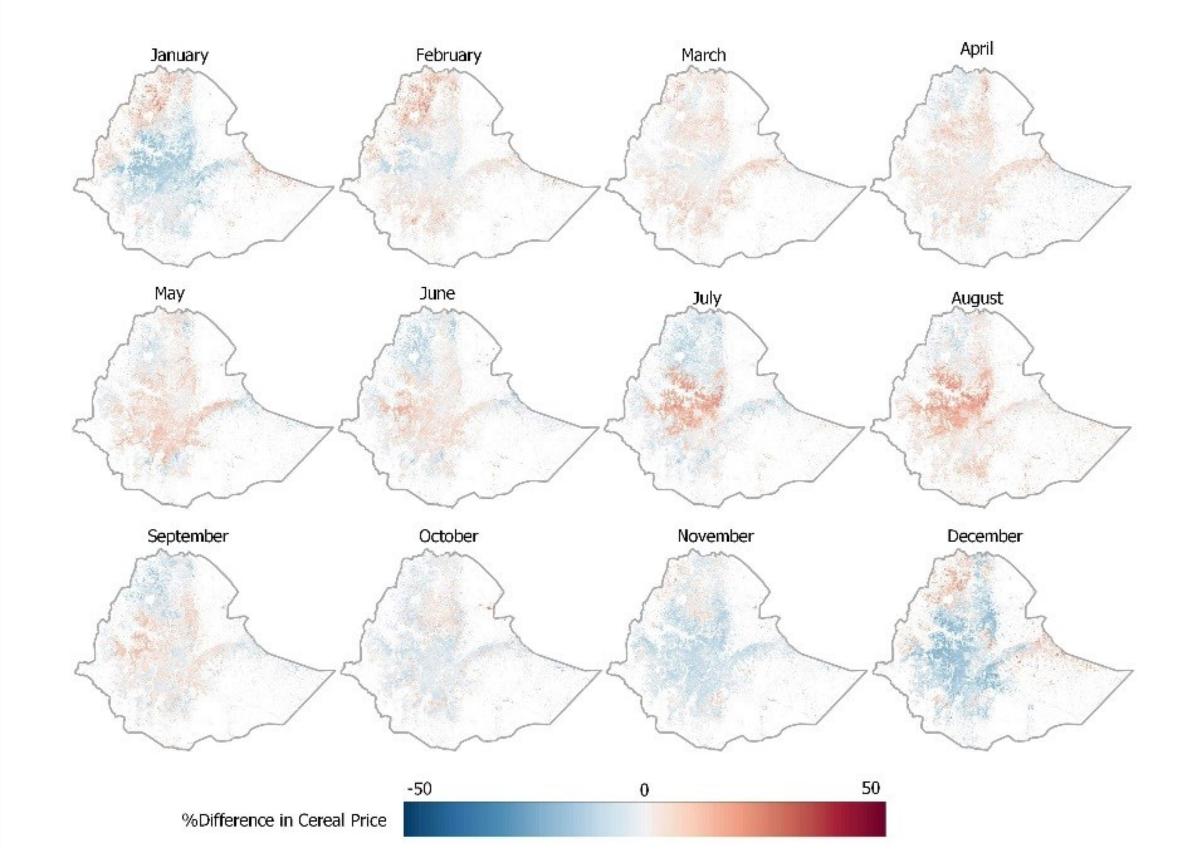


Figure 30. Maps of monthly percentage difference in cereal price in Ethiopia.

7.2. GAM and MLR comparison

A performance comparison was done for the GAM and the multilinear regression (MLR) model. The R^2, RMSE, and the nRMSE for both methods across both training and validation sets are shown in Table 7.

Table 7. Accuracy metrics of GAM and MLR for validation data.

Metric	GAM	MLR
R^2	0.62	0.35
RMSE	3.99	5.17
nRMSE	14.63%	30.67%

GAM attained an R^2 of 0.62 in the validation set, better than that of the MLR with 0.35. Also, GAM managed an RMSE of 3.99, which meant an nRMSE of ~15%, both lower than the RMSE and nRMSE of the MLR which were 5.17 and ~30%, respectively. Overall, the GAM outperformed the MLR in all three performance metrics. This is likely because the relationship between the predictor variables and the FAI were expected to be nonlinear in nature, hence the worse performance of the MLR.